

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:39:47 ; Search time 48 Seconds

(without alignments)
3025.613 Million cell updates/sec

Title: US-09-987-701-2

Perfect score: 2726

Sequence: 1 MSISSEWFLVRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2711	99.4	514	4 AAB95225	Aab95225 Human pro
2	2708	99.3	514	6 ABC07190	Abc07190 Human p53
3	2437	89.4	577	7 ADD14051	Add14051 Human src
4	2340	85.8	542	5 ABP41760	Abp41760 Human ova
5	2047	75.1	395	5 ABP51424	Abp51424 Human MDD
6	1907.5	70.0	584	4 ABG21351	Abg21351 Novel hum
7	1904.5	69.9	700	4 ABB60376	Abb60376 Drosophil
8	1137	41.7	208	4 AAO04385	Aao04385 Human pol
9	819	30.0	167	4 ABG21350	Abg21350 Novel hum
10	499.5	16.3	535	6 ABR52980	Abr52980 Protein s
11	374.5	13.7	414	6 ADA33321	Ada33321 Human int
12	370.5	13.6	481	4 ABB59486	Abb59486 Drosophil.
13	370	13.6	479	3 AAY79678	Aay79678 Drosophil
14	368	13.5	411	4 ABB62260	Abb62260 Drosophil
15	357.5	13.5	485	4 AAB68284	Aab68284 Amino aci
16	356.5	13.4	485	4 AAB68516	Aab68516 Human GTP
17	356.5	13.4	485	4 AAB92844	Aab92844 Human pro
18	356.5	13.4	485	5 ABB97306	Abb97306 Novel hum
19	355.5	13.4	515	6 ABR53774	Abr53774 Protein s
20	354.5	13.4	514	2 AAR55881	Aar55881 WD-40 dom
21	359	13.2	352	3 AAG14893	Aag14893 Arabidops
22	359	13.2	352	3 AAG48119	Aag48119 Arabidops
23	359	13.2	352	7 ADB95040	Adb95040 A. thalia
24	359	13.2	484	4 AAB68282	Aab68282 Amino aci
25	356	13.1	349	3 AAG48120	Aag48120 Arabidops

26	356	13.1	349	3 AAG14894	Aag14894 Arabidops
27	348	12.8	339	2 AAW55957	Aaw55957 Human cia
28	348	12.8	339	2 AAW58888	Aaw58888 Human lip
29	348	12.8	339	7 ADE59693	Ades9693 Human pro
30	348	12.8	339	7 ADE59705	Ades9705 Human pro
31	348	12.8	339	7 ADE59697	Ades9697 Human pro
32	348	12.8	339	7 ADE59709	Ades9709 Human pro
33	348	12.8	339	7 ADE83378	Ades8378 Human pro
34	348	12.8	339	7 ADE89701	Ades9701 Human pro
35	348	12.8	339	7 ADE59713	Ades9713 Human pro
36	348	12.8	339	7 ADE57861	Ades7861 Human pro
37	348	12.8	450	6 ABB99407	Abb99407 Amino aci
38	348	12.8	451	6 ADA21145	Ada21145 Human sec
39	348	12.8	478	4 AAM93784	Aam93784 Human pol
40	348	12.8	478	6 ABB99402	Abb99402 Amino aci
41	348	12.8	521	4 ABB10141	Abb10141 Human cdn
42	348	12.8	521	5 ABE66728	Abp66728 Human pol
43	346.5	12.7	409	7 ADE54235	Ades4235 Human pro
44	344.5	12.6	409	2 AAR70002	Aar70002 OPDE 45 k
45	344.5	12.6	409	7 ADE58488	Ades8488 Rat Prote

ALIGNMENTS

RESULT 1
AAB95225
ID AAB95225 standard; protein; 514 AA.
XX AAB95225;
AC
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17352.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
OS
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-003000253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
PA
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
PS Claim 8; SEQ ID NO 17352; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH0166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SQ Sequence 514 AA;

Query Match 99.4%; Score 2711; DB 4; Length 514;
 Best Local Similarity 99.4%; Pred. No. 1.1e-242;
 Matches 511; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSISDEVNFLVYRYLQESGFHSAFTFGIKSHISQSNINGALVPPAALISIIQGLQVY 60
 DB 1 MSISDEVNFLVYRYLQESGFHSAFTFGIESHISQSNINGALVPPAALISIIQGLQVY 60
 QY 61 EAEVSI NEDGTLFDGRPIESLSLIDAVMPDVVQYRDKLAQQQAAAAAASQ 120
 DB 61 EAEVSI NEDGTLFDGRPIESLSLIDAVMPDVVQYRDKLAQQQAAAAAASQ 120
 QY 121 QGSAKNGENTANGENGAAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFICAWNPV 180
 DB 121 QGSAKNGENTANGENGAAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFICAWNPV 180
 QY 181 SDLLASGSGDSTARVNLSENSTSGTQVLRHCIRREGQDVPNSKDVTSLDWNSGTL 240
 DB 181 SDLLASGSGDSTARVNLSENSTSGTQVLRHCIRREGQDVPNSKDVTSLDWNSGTL 240
 QY 241 ATGSDYGFARITWKDGNLSTLQGHKGFIPALKWKNKGNFSLSGVDTKTIIDWHTGEA 300
 DB 241 ATGSDYGFARITWKDGNLSTLQGHKGFIPALKWKNKGNFSLSGVDTKTIIDWHTGEA 300
 QY 301 KOQFFPHAPALVDWQNNITFASCTDMCHVCKLGDRIKTFQGHTEVNAIKWDPT 360
 DB 301 KOQFFPHAPALVDWQNNITFASCTDMCHVCKLGDRIKTFQGHTEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGFTNNPNANMLASGF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGFTNNPNANMLASGF 420
 QY 421 DSTVRLWDVDRGICHTLTKQEPYVSFAFSDGRLYASGDFDKCVHWTQTALVHSY 480
 DB 421 DSTVRLWDVDRGICHTLTKQEPYVSFAFSDGRLYASGDFDKCVHWTQTALVHSY 480
 QY 481 RTGGIFEVCMNAAGDKVGASDGSVCVLDLRK 514
 DB 481 RTGGIFEVCMNAAGDKVGASDGSVCVLDLRK 514

RESULT 2
 ABO07190
 ID ABO07190 standard; protein; 514 AA.

XX ABO07190;
 AC ABO07190;
 XX 13-AUG-2003 (first entry)
 DE Human p53 modifying protein, SEQ ID 150.
 KW Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
 KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
 KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
 KW apoptotic disorder; cell proliferation disorder.

OS Homo sapiens.
 XX WO200299122-A1.
 PN
 XX 12-DEC-2002.
 PD
 XX 03-JUN-2002; 2002WO-US017382.
 PF
 XX 05-JUN-2001; 2001US-0296076P.
 PR
 XX 10-OCT-2001; 2001US-0328605P.
 PR
 XX 15-FEB-2002; 2002US-0357253P.
 XX (EXEL-) EXELIXIS INC.
 XX Friedman L, Plozman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
 PI
 XX WPI: 2003-156859/15.
 DR N-PSDB; AC013365.
 DR
 XX Identifying modulators of the p53 pathway for use in treating apoptotic
 PT or cell proliferation disorders, comprises screening for agents that
 PT modulate activity of a human ortholog of genes that modify the p53
 PT pathway in Drosophila.
 XX
 PS Example 2; Page 458-459; 678pp; English.
 XX The invention relates to identifying (M1) a candidate p53 pathway
 CC modulating agent, by contacting an assay system comprising a purified HM
 CC polypeptide (human orthologue of genes that modify the p53 pathway in
 CC Drosophila) or nucleic acid with a test agent under conditions, where but
 CC for the presence of the test agent, the system provides a reference
 CC activity, and detecting a test agent-biased activity of the assay system.
 CC Also included are modulating (M2) a p53 pathway of a cell (comprising
 CC contacting a cell defective in p53 function with a candidate modulator
 CC that specifically binds to a HM polypeptide comprising an HM amino acid
 CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
 CC in a mammalian cell (comprising contacting the cell with an agent that
 CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
 CC a disease in a patient (comprising: (a) obtaining a biological sample
 CC from the patient; (b) contacting the sample with a probe for HM
 CC expression; (c) comparing the results with a control; and (d) determining
 CC whether the comparison indicates a likelihood disease). (M1) is useful
 CC for identifying modulators of the p53 pathway. A probe for HM expression
 CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
 CC in a patient, where the cancer has greater than 25 % expression level.
 CC Modulators identified by (M1) are useful in a variety of diagnostic and
 CC therapeutic applications, where disease or disorder prognosis is related
 CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
 CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
 CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
 CC the p53 function of the cell, so that the cell undergoes normal
 CC proliferation or progression through the cell cycle. (M2) and (M3) are
 CC also useful for treating defects in the p53 pathway such as angiogenic,
 CC apoptotic or cell proliferation disorders. The present sequence
 CC represents a human p53 pathway modifying protein
 XX
 SQ Sequence 514 AA;

Query Match 99.3%; Score 2708; DB 6; Length 514;
 Best Local Similarity 99.4%; Pred. No. 2.2e-242;
 Matches 511; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSISDEVNFLVYRYLQESGFHSAFTFGIKSHISQSNINGALVPPAALISIIQGLQVY 60
 DB 1 MSISDEVNFLVYRYLQESGFHSAFTFGIESHISQSNINGALVPPAALISIIQGLQVY 60
 QY 61 EAEVSI NEDGTLFDGRPIESLSLIDAVMPDVVQYRDKLAQQQAAAAAASQ 120
 DB 61 EAEVSI NEDGTLFDGRPIESLSLIDAVMPDVVQYRDKLAQQQAAAAAASQ 120
 QY 121 QGSAKNGENTANGENGAAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFICAWNPV 180
 DB 121 QGSAKNGENTANGENGAAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFICAWNPV 180

QY 181 SLLASGSGSTARIWNLSENSTSGSTOLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 DB 181 SLLASGSGSTARIWNLSENSTSGSTOLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 QY 241 ATGSYDGFARITWKDGNLSTGLGKHGPIFALKWNNKGNFILSAGVDKTTIIWDAHTGEA 300
 DB 241 ATGSYDGFARITWKDGNLSTGLGKHGPIFALKWNNKGNFILSAGVDKTTIIWDAHTGEA 300
 QY 301 KQPPFHSAPALVDWQSNFTASCSTDMCIHVCKLQGDPIKTFQGHTEVNAIKWDPT 360
 DB 301 KQPPFHSAPALVDWQSNFTASCSTDMCIHVCKLQGDPIKTFQGHTEVNAIKWDPT 360
 QY 361 GNILASGSDMTLKIWSMKQDNCVHDLQOHNKEIYITIKNSPTGPGTNNPNANMLASAF 420
 DB 361 GNILASGSDMTLKIWSMKQDNCVHDLQOHNKEIYITIKNSPTGPGTNNPNANMLASAF 420
 QY 421 DSTVRLWDVRGICHTLTKHOEPVYVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSY 480
 DB 421 DSTVRLWDVRGICHTLTKHOEPVYVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSY 480
 QY 481 RGTGGEVFNWNAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RGTGGEVFNWNAAGDKVGASASDGSVCVLDLRK 514

RESULT 3

ID ADD14051 standard; protein; 577 AA.

AC ADD14051;

XX 01-JAN-2004 (first entry)

XX Human src biomarker polypeptide SEQ ID NO:240.

XX predictor set; protein tyrosine kinase activity modulator;
 KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
 KW gene therapy; drug sensitivity; genetic profile; cancer; human.

XX Homo sapiens.

XX WO2003062395-A2.

XX 31-JUL-2003.

XX 17-JAN-2003; 2003WO-US001981.

XX 18-JAN-2002; 2002US-0350061P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Huang F, Fairchild CR, Lee FY, Shaw P;

XX WPI; 2003-636735/60.

XX N-PSDB; ADD14646.

XX New polynucleotides and polypeptides for predicting the activity of
 PT compounds that interact with protein tyrosine kinases and/or protein
 PT tyrosine kinase pathways.

XX Claim 10; SEQ ID NO 240; 139pp; English.

XX The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and
 CC correlating the expression of the markers to the compound's ability to
 CC modulate the activity of the cells; (2) a plurality of cell lines for
 CC identifying polynucleotides and polypeptides whose expression levels

CC correlate with compound sensitivity or resistance of cells associated
 CC with a disease state; and (3) identifying polynucleotides and
 CC polypeptides that predict compound sensitivity or resistance of cells
 CC associated with a disease state, comprising subjecting the plurality of
 CC cell lines to one or more compounds, analysing the expression pattern of
 CC a microarray of polynucleotides or polypeptides, and selecting
 CC polynucleotides or polypeptides that predict the sensitivity or
 CC resistance of cells associated with a disease state by using the
 CC expression pattern of the microarray. The polynucleotides and
 CC polypeptides have cytostatic activities, and can be used in gene therapy.
 CC The polynucleotides and polypeptides are useful in predicting the
 CC activity of compounds that interact with protein tyrosine kinases and/or
 CC protein tyrosine kinase pathways. These may be used in determining drug
 CC sensitivity in patients to allow the development of individualized
 CC genetic profiles which aid in treating diseases and disorders (e.g.
 CC cancer) based on patient response at a molecular level. The present
 CC sequence is used in the exemplification of the present invention.

XX Sequence 577 AA;

Query Match 89.4%; Score 2437; DB 7; Length 577;

Best Local Similarity 86.0%; Pred. No. 4e-217;

Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

QY 1 MSISDEVNFLVRYLQBSGFSHSAFTGKSHISQSNINGALVPPAALISIIQGLQV 60
 DB 52 MSITSDEVNFLVRYLQBSGFSHSAFTGKSHISQSNINGALVPPAALISIIQGLQV 111
 QY 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQQAAAAA--- 117
 DB 112 EAEISNEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQQAAAAAATAA 171
 QY 118 -----ASQQSAXNGENTANGENGAAHTIANNHTDMMMEVDGVEIIPNNKAVVLRG 167
 DB 172 ATAATTAGVSHQNPSPKNREATVNGEENRAHSV-NNHAKPMEIDGEVEIPSSKATVLRG 230
 QY 168 HSEVFIQAWNPVSDLLASGSGDSTARWNLSENSTSGTOLVLRHCIREGGQDVPSNKD 227
 DB 231 HSEVFIQAWNPVSDLLASGSGDSTARWNLSENSTSGTOLVLRHCIREGGQDVPSNKD 290
 QY 228 VTSLDWNSGTLTATGSDYDGFARIWTKDGNLSTGLGKHGPIFALKWNNKGNFILSAGVD 287
 DB 291 VTSLDWNTGTLTATGSDYDGFARIWTKDGNLSTGLGKHGPIFALKWNNKGNFILSAGVD 350
 QY 288 KTTIWDHAHTGEAKQFPFHSAPALVDWQSNFTASCSTDMCIHVCKLQGDPIKTFQ 347
 DB 351 KTTIWDHAHTGEAKQFPFHSAPALVDWQSNFTASCSTDMCIHVCKLQGDPIKTFQ 410
 QY 348 HTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYITIKNSPTGPGTN 407
 DB 411 HTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYITIKNSPTGPGTN 470
 QY 408 NPNANMLASAFDSTVRLWDVRGICHTLTKHOEPVYVAFSPDGRYLAGSFDKCVH 467
 DB 471 NPNSNIMLASAFDSTVRLWDVRGICHTLTKHOEPVYVAFSPDGRYLAGSFDKCVH 530
 QY 468 IWNTOGALVHSYRGTTGTFEVCWNAAGDKVGASASDGSVCVLDLRK 514
 DB 531 IWNTOGALVHSYRGTTGTFEVCWNAAGDKVGASASDGSVCVLDLRK 577

RESULT 4

ABP41760

ID ABP41760 standard; protein; 542 AA.

XX ABP41760;

XX 22-AUG-2002 (first entry)

XX Human ovarian antigen HNOKM38, SEQ ID NO:2892.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;

KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 OS Homo sapiens.
 XX
 XX WO200200677-A1.
 PN
 XX
 XX 03-JAN-2002.
 PD
 XX
 XX 07-JUN-2001; 2001WO-US018569.
 PF
 XX
 XX 07-JUN-2000; 2000US-0209467P.
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Birse CE, Rosen CA;
 PI
 XX
 XX WPI. 2002-147878/19.
 DR
 XX
 XX N-PSDB; ABQ54837.

Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 cancer), immune disorders, cardiovascular disorders and neurological
 diseases.

Claim 11; SEQ ID NO 2892; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-
 ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 encompasses polypeptides 90% identical and polynucleotides 95% identical
 to the sequences of the invention. The invention additionally relates to
 recombinant vectors and host cells comprising human ovarian antigen
 polynucleotides, antibodies against human ovarian antigens, and the use
 of ovarian antigen polynucleotides and polypeptides in diagnosing,
 treating, prognosing or preventing various ovary and/or breast-related
 disorders. Such conditions include ovarian cancer and breast cancer, and
 metastatic tumours of ovarian or breast origin, reproductive system
 disorders (e.g., infertility, disorders of pregnancy, anovulation,
 polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 vaginitis), immune disorders (e.g., congenital and acquired
 immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 blood-related disorders (e.g., anaemia), cardiovascular disorders,
 respiratory disorders, neurological disorders, gastrointestinal disorders
 and urinary system disorders. Ovarian antigen polypeptides and
 polynucleotides may also be used in screening for compounds which
 modulate ovarian antigen expression or activity. The polynucleotides may
 further be used for gene therapy, chromosome mapping, in the
 identification of individuals and in forensic analysis, and the
 polypeptides may be used as food additives or to prepare antibodies
 useful in disease diagnosis, drug targeting and phenotyping. The present
 sequence represents a human ovarian antigen of the invention. Note: The
 sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences

Sequence 542 AA;

Query Match 85.8%; Score 2340; DB 5; Length 542;
 Best Local Similarity 85.4%; Fred. No. 3.8e-206;
 Matches 434; Conservative 32; Mismatches 28; Indels 14; Gaps 2;
 20 GFSHSAFTFGKSHISQSNINGALVPPAALISIIQKGLQYVEAEVSINEDGTLFDGRPIE 79
 36 GFSHSAFTFXLESHISQSNINGTLVPPAALISIIQKGLQYVEAEISINEDGTVFDGRPIE 95

QY 80 SLSLIDAVMPDVVQTRQAYRDKLAQQAAAAA-----ASQCSAKN 126
 DB 96 SLSLIDAVMPDVVQTRQAFREKLAQQASAAAAAATAAATAATTTAGVSHQNSKN 155
 QY 127 GENTANGENGAAHTTANNHTDMVDGVEIPPNKAVLVRGHSEVFCAMNPYSILLAS 186
 DB 156 REATVNGEENRAHSV-NNHAKPMEIDGVEIPESSKATVLRGHSEVFCAMNPYSDLLAS 214
 QY 187 GSGDSTARLWNLSENSTGSLVLRHCIREGQDVPNSKDVTSIDMNSEGLLATGSYD 246
 DB 215 GSGDSTARLWNLSENNGSTQLVLRHCIREGHDVPSNKDVTSIDMNTNGLLATGSYD 274
 QY 247 GFARIWTKDGNLASTLGOHKGPFFALKKNKKNFILSAGVDKTTIWDHAHTGEAKQPPF 306
 DB 275 GFARIWTEGDNLASTLGOHKGPFFALKKNRKNGYILSAGVDKTTIWDHAHTGEAKQPPF 334
 QY 307 HSAPALDWDQSNNTFASCSITDWCIVHCKIGDRPIKTFQGHNEVNNAIKWDPCTGNLLAS 366
 DB 335 HSAPALDWDQNNITFASCSITDWCIVHCKIGDRPVKTFQGHNEVNNAIKWDPGKLLAS 394
 QY 367 CSDDMTLKIWSMKQDNCVHDLQOHNKETIYTIKWSPTGPTNNPNANMLASAFDSTVRL 426
 DB 395 CSDDMTLKIWSMKQEVCIHDLQAHNKETIYTIKWSPTGPTSNPNINMLASAFDSTVRL 454
 QY 427 WVDVDRGICHTLTIKHQPYPYVAFSPDGRYLASGSGFDKCVHNTOTGALVHSYRGTTGI 486
 DB 455 WDIERGVCTHTLTIKHQPYPYVAFSPDGRYLASGSGFDKCVHNTOTGALVHSYRGTTGI 514
 QY 487 FEVCNNAAGDKVGASASDGSVCVLDLRK 514
 DB 515 FEVCNNAAGDKVGASASDGSVCVLDLRK 542
 RESULT 5
 ABP51424
 ID ABP51424 standard; protein; 395 AA.
 XX
 AC ABP51424;
 XX
 DT 03-SEP-2002 (first entry)
 XX
 DE Human MDDT SEQ ID NO 446.
 XX
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiatherosclerotic;
 KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;
 KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antigout;
 KW neuroprotective; antirheumatic; antiarthritic.
 XX
 OS Homo sapiens.
 XX
 XX WO200240715-A2.
 PN
 XX 23-MAY-2002.
 PD
 XX
 XX 06-SEP-2001; 2001WO-US02762B.
 PF
 XX
 XX 05-SEP-2000; 2000US-0229747P.
 PR
 XX 05-SEP-2000; 2000US-0229748P.
 PR
 XX 05-SEP-2000; 2000US-0229749P.
 PR
 XX 05-SEP-2000; 2000US-0229750P.
 PR
 XX 05-SEP-2000; 2000US-0229751P.
 PR
 XX 05-SEP-2000; 2000US-0230583P.
 PR
 XX 05-SEP-2000; 2000US-0230505P.
 PR
 XX 05-SEP-2000; 2000US-0230514P.
 PR
 XX 05-SEP-2000; 2000US-0230515P.
 PR
 XX 05-SEP-2000; 2000US-0230517P.
 PR
 XX 05-SEP-2000; 2000US-0230518P.
 PR
 XX 05-SEP-2000; 2000US-0230519P.
 PR
 XX 05-SEP-2000; 2000US-0230595P.
 PR
 XX 05-SEP-2000; 2000US-0230597P.


```
PR 06-SEP-2000; 2000US-0230598P.
PR 06-SEP-2000; 2000US-0230599P.
PR 06-SEP-2000; 2000US-0230610P.
PR 06-SEP-2000; 2000US-0230612P.
PR 06-SEP-2000; 2000US-0230663P.
PR 06-SEP-2000; 2000US-0230988P.
PR 06-SEP-2000; 2000US-0230989P.
PR 07-SEP-2000; 2000US-0230951P.
PR 07-SEP-2000; 2000US-0231163P.
PR 07-SEP-2000; 2000US-0231167P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MG;
XX Hillman JL, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE;
XX Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B;
XX Roseberry AM, Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V;
XX Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Iman RR;
XX WPI; 2002-527544/56.
XX N-PSDB; ABQ72641.
XX
XX Novel human disease detection and treatment polypeptide, useful in
XX diagnosis, prevention or treatment of cell proliferative disorders e.g.
XX arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.
XX AIDS.
XX
XX Claim 14; Page 578; 618pp; English.
XX
XX The invention relates to an isolated human disease detection and
XX treatment (MDPT) polypeptide (I) selected from a polypeptide having a
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the
XX specification, a naturally occurring polypeptide comprising a sequence
XX having at least 90% identity to (I) or a biologically active or
XX immunogenic fragment of (I). (I) is useful for screening a compound for
XX effectiveness as an agonist or antagonist, for screening a compound that
XX specifically binds (I) or modulates the activity of (I), and for
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.
XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
XX screening a compound for effectiveness in altering expression of a target
XX polynucleotide comprising. Oligonucleotides and antibodies are useful for
XX detecting MDPT in a sample or for assessing toxicity of a test compound,
XX in a diagnostic test for a condition or a disease associated with the
XX expression of MDPT in a biological sample, for detecting (I) in a sample,
XX and for purifying (I) from a sample. A composition comprising (I), an
XX agonist or antagonist is useful for treating a disease or condition
XX associated with decreased or increased expression of functional MDPT. (I)
XX or (II) are useful for diagnosing, treating or preventing disorders
XX associated with aberrant expression of MDPT, where the disorders are
XX selected from a cell proliferative disorder such as arteriosclerosis,
XX cirrhosis, hepatitis, psoriasis, and cancer and an
XX autoimmune/inflammatory disorder such as AIDS, Addison's disease,
XX allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
XX rheumatoid arthritis. (II) are useful for creating knockin humanised
XX animals or transgenic animals to model human diseases, in somatic or
XX germline gene therapy, to generate a transcript image of a tissue or cell
XX type, for detecting differences in the chromosomal location due to
XX translocation or inversion among normal, carrier or affected individuals
XX and as hybridisation probes for mapping naturally occurring genomic
XX sequences
XX
XX Sequence 395 AA;
XX
XX Query Match
XX Best Local Similarity 75.1%; Score 2047; DB 5; Length 395;
XX Matches 389; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 MSISDEVNFLVRYLQESGFSHSAFTGKISHISQSNINGALVPPAALISIIQKGLQYV 60
XX DB 5 MSISDEVNFLVRYLQESGFSHSAFTGKISHISQSNINGALVPPAALISIIQKGLQYV 64
XX
XX 61 EAEVINEEDGTLFCGRPTESLSLDVMPDVVQTEQAYRDKLACQAAAAAAAASQ 120
XX DB 65 EAEVINEEDGTLFCGRPTESLSLDVMPDVVQTEQAYRDKLACQAAAAAAAASQ 124
```

```
QY 121 QGSAKNGENTANGENGCAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFCANPV 180
DB 125 QGSAKNGENTANGENGCAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFCANPV 184
QY 181 SDLLASGSGDSTARINWLSNSTSGSTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
DB 185 SDLLASGSGDSTARINWLSNSTSGSTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 244
QY 241 ATGSYDGFARINTKDNGLASTLQHKGPPIPAKKNKGNFILSAGVDKTTIIIDAHTGEA 300
DB 245 ATGSYDGFARINTKDNGLASTLQHKGPPIPAKKNKGNFILSAGVDKTTIIIDAHTGEA 304
QY 301 KQFPFPHSAFALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
DB 305 KQFPFPHSAFALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 364
QY 361 GNLLASCSDDMTLKIWSMKODNCVHDLQAHN 391
DB 365 GNLLASCSDDMTLKIWSMKODNCVHDLQAHN 395
```

RESULT 6

```
ABG21351
ID ABG21351 standard; protein; 584 AA.
XX AC ABG21351;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #21342.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX OS Homo sapiens.
XX FN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PA 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS85538.
```

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID NO 51710; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in

RESULT 7	
ABB60376	
ID ABB60376	standard; protein; 700 AA.
XX	
XX	
AC	ABB60376;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 7920.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
FN	WO200171042-A2.
XX	

QY 353 NAIKWDPTGNLLASCSDDMTLKIMSKQDNCVHDLQHNKEIYTIKWSPTGPGTNNPNAN 412
 DB 538 NAIKWCPOGQULASCSDDMTLKIMSMNRDRCHDLQAHSEKIYTIKWSPTGPGTNNPNTN 597
 QY 413 LMLASASDSTVRLWDVDRGICIIHTLTKHQBPVYSVAFSPDGRVLAGSPDKCVHIWNTQ 472
 DB 598 LILASASDSTVRLWDVDRGICIIHTLTKHQBPVYSVAFSPDGRVLAGSPDKCVHIWNTQ 657
 QY 473 TGALVHSYRGTGGIFEVWCNAGDKVGSASDGSVCVLDLRK 514
 DB 658 TQGLVHSYKGTGGIFEVWCNAGDKVGSASDGSVCVLDLRK 699

RESULT 8
 ID AAO04385
 AC AAO04385;
 DT 06-NOV-2001 (first entry)
 DE Human polypeptide SEQ ID NO 18277.
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX Homo sapiens.
 OS
 XX WO200164835-A2.
 PN
 XX 07-SEP-2001.
 PD
 XX 26-FEB-2001; 2001WO-US004927.
 PF
 XX 28-FEB-2000; 2000US-0051526.
 PR
 XX 18-MAY-2000; 2000US-00577409.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI
 XX WPI; 2001-514838/56.
 DR
 XX N-PSDB; AAI84316.
 DR
 XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.
 PT
 XX Claim 20; SEQ ID NO 18277; 1399pp + Sequence Listing; English.
 PS
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 208 AA;
 Query Match 41.7%; Score 1137; DB 4; Length 208;
 Best Local Similarity 99.5%; Pred. No. 5.5e-97;
 Matches 206; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 213 HCIREGGQDVPSKDVTSLDWNSRGTLATGSDGFARITWKDGNLASTLGQHKGPITAL 272
 |||||||

DB 1 HCIREGGQDVPSKDVTSLDWNSRGTLATGSDGFARITWKDGNLASTLGQHKGPITAL 60
 QY 273 KNNKGNFILSAGVDKTTIIWDAHTGEAKQOPFFHSAPALDWDQSNNTFASGSTDWCIH 332
 DB 61 KNNKGNFILSAGVDKTTIIWDAHTGEAKQOPFFHSAPALDWDQSNNTFASGSTDWCIH 120
 QY 333 VCKLGQDRPIKTFQGHTEVNAIKWDPGTGNLLASCSDDMTLKIMSKQDNCVHDLQHNK 392
 DB 121 VCKLGQDRPIKTFQGHTEVNAIKWDPGTGNLLASCSDDMTLKIMSKQDNCVHDLQHNK 180
 QY 393 EYTIKWSPTGPGTNNPNANLMIASAS 419
 DB 181 EYTIKWSPTGPGTNNPNANLMIASAS 207

RESULT 9
 ID ABG21350
 AC ABG21350;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #21341.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US008631.
 PF
 XX 31-MAR-2000; 2000US-00540217.
 PR
 XX 23-AUG-2000; 2000US-00649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR
 XX N-PSDB; AAS85537.
 DR
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 XX Claim 20; SEQ ID NO 51709; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (II) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 167 AA;

Query Match 30.0%; Score 819; DB 4; Length 167;
Best Local Similarity 86.2%; Pred. No. 1.4e-67;
Matches 144; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 250 RIWTKDGNLSTGLGKGFIFALKWKKGNFILSAGVDKTTIWDHAHGEAKQFPFSA 309
DB 1 RIWTEGNLAIITLCKHKGPIFALKWKKGNVLSAGVDKTTIWDHAHGEAKQFPFSA 60

QY 310 PALVDWQSNNTFASCSTDMCIHVCKLQDRIPIKTFQGHTEVNAIKWDPTGNLLASCSD 369
DB 61 PALVDWQNNMTFASCSTDMCIHVCKLQDRIPIKTFQGHTEVNAIKWDPTGNLLASCSD 120

QY 370 DMTLKWSMKQDNCVHDLQHNKEIYTIKWSPTGPTGNPNANLMLA 416
DB 121 DMTLKWSMKQEVCIHDLQHNKEIYTIKWSPTGPTGNPNANLMLA 167

RESULT 10
ID ABR52980 standard; protein; 535 AA.
XX ABR52980;
XX
XX
DT 20-JUN-2003 (first entry)
XX
DE Protein sequence #SEQ ID 825.
XX
XX Multiprotein complex; eukaryote; drug target; diagnosis.
XX
XX Saccharomyces cerevisiae.
XX
XX EP1258494-A1.
XX
XX 20-NOV-2002.
XX
XX 20-DEC-2001; 2001EP-00130253.
XX
XX 15-MAY-2001; 2001EP-00111774.
XX
XX (CELL-) CELLZONE AG.
XX
XX Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
PI Marzioch M, Schultz JD, Superti-Furga GD;
XX
XX N-PSDB; ACC61022.
XX
XX WPI; 2003-250078/25.
XX
XX
XX New isolated protein complexes useful for diagnosing a disease or
PT disorder, or as a target for an active agent of a pharmaceutical,
PT preferably a drug target in the treatment or prevention of disease or
PT disorder.
XX
XX Disclosure; SEQ ID NO 825; 17pp + Sequence Listing; English.
XX
XX The invention relates to multiprotein complexes from eukaryotes. Proteins
XX of the invention and DNA sequences encoding them are given in records
XX of the invention and ACC60610-ACC61944 respectively. The complexes are
XX obtainable by using a protein as a bait and isolating the set of proteins
XX which is attached thereto from cells. Such protein complexes may comprise
XX up to 30 distinct proteins. Protein complexes of the invention are useful
XX for diagnosing a disease or disorder, or as a target for an active agent
XX of a pharmaceutical, preferably a drug target in the treatment or
XX prevention of a disease or disorder. Note: The sequence data for this
XX patent is not represented in the printed specification, but is based on
XX sequence information supplied by the European Patent Office. The complete
XX document is available on CD-ROM
XX
XX Sequence 535 AA;

Query Match 18.3%; Score 499.5; DB 6; Length 535;
Best Local Similarity 24.9%; Pred. No. 4.3e-37;
Matches 142; Conservative 100; Mismatches 232; Indels 97; Gaps 15;

QY 1 MSITSEVNFLVRYLQESGFSHA-----FTFGIKSHISQSNINGALVPPAAL 49
DB 1 MSITSEVNFLVRYLQESGFSHA-----FTFGIKSHISQSNINGALVPPAAL 50

QY 50 ISIIQGLGVYAEVSNEDG---TLFQGRPIESLSDAVMPDVVQVQOAYRKLAAQ 106
DB 51 VNLVQSGIITSELMVDSKGDISALNEHLSDFNLVQALQID-----KEKFEI 101

QY 107 QAAAAAASQASAKNGENTANGENGAAHTIANHNTDMVEVDGVEIPPNKAVLVR 166
DB 102 SSGRTFLTESNKGAGDGAETVERTOEDTNSIDSD--DLDFVKI--LKEIV-- 155

QY 167 GHESEVFI CAMNPVSD--LLASGSDSTARIMWLSNENSTSGSTQ-----LVLRHCIREGG 219
DB 156 -KLDNIVSSTWNPDESILAYGEKNSVARLARIVETDQEGKYYKLTITIAELRHPFALSA 214

QY 220 QDVPSNKDVTSLDWNSEGLTLAGSVDFGPARITWTKDGNLSTGLGKGFIPALKWKKGN 279
DB 215 SSGKTTNQVTCIAWSHDGNSIVTGVENGELRWKGTALLNVLNFRAPIVSVKANKDGT 274

QY 280 FILSAGVDKTTIWDHAHGEAKQFPF-----HSAP--ALDVMQSNNTFAS 324
DB 275 HIISMDVENVTILNVIISGVNQHPELKETGSSNAENHSGDGLGVDEWVDDDKFVI 334

QY 325 CSTDMCIHVCKLQDRIPIKTFQGHTEVNAIKWDPTGNLLASCSDMTLKWSMKQDNCV 384
DB 335 PGPKGAIFYQITKTEPTCKLIHHGPISVLEBFDNKLKLSASDDGTURIWHGNGNSQ 394

QY 385 HDLQHNKEIYTIKWSPTGPTGNPNANLMLASASDFSTVRLWDVDRGICHTLTKHOEP 444
DB 395 NCFYHGSQIVSNVW-----GDDKVISCSWDGSGVRLWSLKNQNTLLALSIVDGV 444

QY 445 VYSVAFSPDGRYLASGSPKCVHIW-----NTQTGAL-----VHSVR 481
DB 445 IFAGRISQDQKYAVAFMDQGVNVYDLKLNKSKRSYLNCRDGIILNPLPIPLVASYQSSQ 504

QY 482 GTGGEFECVWNAAGDKVGA--SASDGSVCVL 510
DB 505 DNDYIFDLSWNCAGNKISVAYSLOEGSVVAI 535

RESULT 11
ID ADA13321 standard; protein; 414 AA.
XX ADA13321;
XX
XX 06-NOV-2003 (first entry)
DT Human intracellular signalling molecule INTSIG-3, SEQ ID NO:3.
DE Human intracellular signalling molecule INTSIG-3, SEQ ID NO:3.
XX
XX Human, intracellular signalling molecule; INTSIG;
XX cell proliferative disorder; cancer; atherosclerosis;
XX autoimmune disorder; inflammatory disorder; infection;
XX neurological disorder; developmental disorder; endocrine disorder;
XX cytostatic; antiarteriosclerotic; nootropic; neuroprotective;
XX cerebroprotective; immunosuppressive; antiinflammatory; gene therapy.
XX Homo sapiens.
XX WO2003031568-A2.
XX
XX 17-APR-2003.
XX
XX 16-AUG-2002; 2002WO-US026322.
XX
XX 17-AUG-2001; 2001US-0313245P.
XX
XX 24-AUG-2001; 2001US-0314751P.
XX
XX 31-AUG-2001; 2001US-0316752P.
XX

Db 168 RLASCKAGSIIPETGQKGRPLSGHKHINCLAWEPVHRPECKKLASAGDGD 227
 QY 251 IW-TKDGNLASTLQGHKGPFIKLNKKNKNGNFIKAGVDKTIINDAHTGEAKQFPFSA 309
 Db 228 IWDVKGQCLMNIAGHTNAVTAVRGGAG-LIYTSSKDRIVKMWRAADGILCRITFSGHA- 285
 QY 310 PALDWDGNSNTFASCDTMCIHVCKLGDRIK-TFOGH-----TNEV-----NAIK 356
 Db 286 -----HWNN-----IALSTD---YVLTGTFPHKDRSKSHLSLSTEELQESALKRYQAVC 334
 QY 357 WDPTGNLLASDDMTLKIWSKQDNCVHDLQKHKEIYTIKWSPGTGPTNNPNNMLLA 416
 Db 335 PDEVESLV-SCSDDNTLYLRNNQNKVCVERMTGHQNVVNDVKYSPDVK-----LIA 384
 QY 417 SASFSTVRLNDVDRGICHTLTKEQEPVYSVARSFGRYVLSGSKFCVHIWNTQTAL 476
 Db 385 SASFDKSVRLWRASDQGYMATFRGHVQAVYIVAMSADSLIVSGSKSTLKWVSQTKL 444
 QY 477 VHSYRG-TGGIFEVCNNAAGDKVGASD 504
 Db 445 AQELPCHADEVEGVDPADGSRVASGGKD 473

RESULT 13

AA79678
 ID AA79678 standard; protein; 479 AA.

XX AC AA79678;
 XX DT 29-AUG-2000 (first entry)
 XX DE Drosophila Notchless protein.
 XX KW Notchless; Nle Gene; Notch; signalling; neurodegenerative disease;
 XX KW cancer; diagnosis; cytostatic; neuroprotective; therapy.
 XX OS Drosophila melanogaster.
 XX FH Key Location/Qualifiers
 XX FT Domain 27..105
 XX FT /note= "Nle domain"
 XX FT Misc-difference 67
 XX FT /note= "sequence deduced from nucleotide sequence has an
 XX FT additional Lys residue between Lys-67 and Ser-68"
 XX FT Misc-difference 282
 XX FT /note= "encoded by CAA"

WO200026364-A1.

PD 11-MAY-2000.

XX 03-NOV-1999; 99WO-IB001891.

XX 03-NOV-1998; 98GB-00024045.

XX (EURO-) EURO MOLECULAR BIOLOGY LAB.

XX Cohen S, Bouwmeester A, Royet J;

XX WPI; 2000-365613/31.

XX N-PSDB; AAA27739.

PT Novel notchless protein and nucleic acids encoding them useful for
 PT treating and preventing cancer and neurodegenerative diseases.

XX Claim 1; Page 44; 52pp; English.

XX The present sequence is that of Notchless, a novel protein of Drosophila.
 CC Notchless was identified in a screen for dominant modifiers of a Notch
 CC mutant phenotype in the Drosophila wing. The mutant dominantly suppressed
 CC the wing notching phenotype of notchoid mutations, and the Notchless
 CC protein was shown to bind to the cytoplasmic domain of Notch. Notchless

CC modified Notch signalling activity in a variety of Notch-dependent
 CC signalling process in both Drosophila and Xenopus embryos. The Notchless
 CC protein has a novel highly conserved N-terminal domain followed by 9 WD40
 CC repeats. Notchless, and nucleic acids encoding it, can be used in methods
 CC for the diagnosis and therapy of certain diseases, particularly cancer
 CC and neurodegenerative diseases (claimed). A Notchless mutant in a
 CC sensitised Notch genetic background is used in a claimed method for
 CC identifying compounds capable of modifying the levels of expression or
 CC activity of a Notch protein
 XX Sequence 479 AA;

Query Match

Best Local Similarity 13.6%; Score 370; DB 3; Length 479;

Matches 120; Conservative 55; Mismatches 158; Indels 90; Gaps 18;

QY 147 DMMEVDG---DVEIIPNKAVVLR-----GHESEVFICAWNPFVSDLLASGSGDST 192
 Db 74 DLASVTENVIDIVYQPAVFKVRPVYTRCTSSMPGHAEAVVSLNFSPOGAHLASGSDTT 133
 QY 193 ARIWNLSEN-----STSGSTQLVLRHCIR-----EGQD----- 221
 Db 134 VRLWDLNTETPHFTCTGCHKQWVL--CVSWAPDGKRLASGCKAGSIIIWDPETGQCKGRPL 191
 QY 222 VPSNKDVTSLDW-----NSEGTLATGSDYDGFARIW-TKDGNLASTLQGHKGPFIKLNKWN 275
 Db 192 SGHKHINCLAWEPVHRPECKKLASAGDGDRIWVKLGQCLMNIAGHTNAVTAVRWG 251
 QY 276 KKGNFILSAGVDKTTIINDAHTGEAKQFPFHSAPALDWDQSNNTFASCDTMCIHVCK 335
 Db 252 GAG-LIYTSSKDRIVKMWRAADGILCRITFSGHA-----HWNN-----IALSTD---YVLR 298
 QY 336 LGQDRPIK-TFOGH-----TNEV-----NAIKWPTGNLLASGSDMTLKIWSKQDN 382
 Db 299 TGFPHFVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLV-SCSDDNTLYLRNNQNK 357
 QY 383 CVHDLQKHKEIYTIKWSPGTGPTNNPNNANLMLASAFDSTVRLNDVDRGICHTLTKHQ 442
 Db 358 CVERMTGHQNVVNDVKYSPDVK-----LIASAFDKSVRLWRASDQGYMATFRGHV 408
 QY 443 EPVYSVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSYRG-TGGIFEVCNNAAGDKVGAS 501
 Db 409 QAVYIVAMSADSLIVSGSKDSTLKWVSQTKLAQELPCHADEVEGVDPADGSRVASG 468
 QY 502 ASD 504
 Db 469 GKD 471

RESULT 14

ABB62260
 ID ABB62260 standard; protein; 411 AA.

XX AC ABB62260;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 13572.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX WO2001171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-C191637P.

XX PR 11-JUL-2000; 2000US-C0614150.


```

Db 289 WNTWALSTDYALRTGAPEAEASVNPQDIQSLQELKERALSRYNLVRGQGPRLVSGS 348
Qy 327 TDMCIHVCKLGOD-RPIKTFQGHTEVNAIKWDPTGNLLASCDDMTLKIWSMKQDNQVH 385
Db 349 DDFTLFLWSPAEDKKPLTRMTGHQALINQVXFSPDSRIVASASFDKSIKLDGRTGKYL 408
Qy 386 DLOQHNEIYTIKWSPTGPTNNPNANMLASAFDSTVRLWDVDRGICIHITLTKHOEPV 445
Db 409 SLRGHVAAYQIANSAD-----SRLVSGSDSTLKWDVKAQKLANDLPCHADEV 459
Qy 446 YSVAFSPDGRVLAGSGFDKQVHIW 469
Db 460 YAVDWSPDGQORVASGGDKCLRIW 483

```

Search completed: August 9, 2004, 16:45:24
 Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:40:42 ; Search time 8.66667 Seconds
(without alignments)
3088.161 Million cell updates/sec

Title: US-09-987-701-2

Perfect score: 2726

Sequence: 1 MSISDEVNFLVRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2716	99.6	514	1 TBLR_HUMAN	Q9bzk7 homo sapien
2	2437	89.4	526	1 TBLX_HUMAN	O60907 homo sapien
3	2367	86.8	522	1 TBLX_HUMAN	Q9bq87 homo sapien
4	1304	47.8	313	1 TBLX_MOUSE	Q9qxe7 mus musculus
5	502.5	18.4	535	1 SIF2_YEAST	P18262 saccharomyc
6	479	17.6	1526	1 YV46_ANASP	Q8yri1 anabaena sp
7	463.5	17.0	1258	1 YS00_ANASP	Q8yrt2 anabaena sp
8	460	16.9	1356	1 HET1_PODAN	Q00808 podospora a
9	445.5	16.3	1683	1 YL24_ANASP	Q8yvs7 anabaena sp
10	388.5	14.3	1693	1 Y163_SYNY3	Q55583 synecocyst
11	366.5	13.4	485	1 HUS7_HUMAN	Q9nvx2 homo sapien
12	365.5	13.4	485	1 YCW2_YEAST	Q8vej4 mus musculus
13	365.5	13.4	515	1 AAC3_DICDI	P25382 saccharomyc
14	348.5	12.8	437	1 CIA1_HUMAN	P14197 dictyosteli
15	348	12.8	339	1 CIA1_HUMAN	O76071 homo sapien
16	346.5	12.7	409	1 LIS1_HUMAN	P43034 homo sapien
17	345.5	12.7	409	1 LIS1_BOVIN	P43033 bos taurus
18	344.5	12.6	409	1 LIS1_MOUSE	P43035 mus musculus
19	340.5	12.5	361	1 WDS_DROME	Q9v318 drosophila
20	339.5	12.5	334	1 WDR5_HUMAN	Q9ugp9 homo sapien
21	339	12.4	800	1 T2D4_HUMAN	Q15542 homo sapien
22	334	12.3	742	1 PKWA_THECU	P49695 thermomonos
23	329	12.1	798	1 T2D4_YEAST	P38129 saccharomyc
24	321	11.8	376	1 YK14_CABEL	Q17963 caenorhabdi
25	319.5	11.7	609	1 PF20_CHLRE	P93107 chlamydomon
26	317.5	11.6	579	1 SE10_CABEL	Q93794 caenorhabdi
27	310.5	11.4	704	1 T2D4_DROME	P49846 drosophila
28	309	11.3	1191	1 Y143_SYNY3	P74442 synecocyst
29	308	11.3	640	1 MT30_YEAST	P39014 saccharomyc
30	307	11.3	678	1 SCOB_TMENTI	Q00659 emericella
31	304	11.2	605	1 POF1_SCHPO	P87053 schizosacch
32	300	11.0	665	1 LI23_CABEL	Q09990 caenorhabdi
33	294.5	10.8	501	1 YH92_CABEL	Q23256 caenorhabdi

ALIGNMENTS

RESULT 1

ID	TBLR_HUMAN	STANDARD;	PRT;	514 AA.
AC	Q9BZK7; Q9H2I1; Q9H9A1;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Nuclear receptor co-repressor/HDAC3 complex subunit TBLR1 (TBL1-related protein 1).			
GN	TBLR1 OR IRA1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND INTERACTION WITH HDAC3.			
RX	MEDLINE=21930350; PubMed=11931768;			
RA	Zhang J., Kalkum M., Chait B.T., Roeder R.G.;			
RT	"The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2.";			
RL	Mol. Cell 9:611-623(2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Zhang X., Dormady S., Basch R.;			
RT	"Identification of four human cDNAs that are differentially expressed by early hematopoietic progenitors.";			
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,			
RA	Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,			
RA	Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;			
RT	"NED0 human cDNA sequencing project.";			
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.			
CC	-1- SUBUNIT: Interacts with HDAC3.			
CC	-1- SIMILARITY: Contains 1 Lish domain.			
CC	-1- SIMILARITY: Contains 8 WD repeats.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AF314544; AAK0301.1; -			
DR	EMBL; AF268193; AAG44736.1; -			
DR	EMBL; AK022956; BAB14331.1; -			
DR	InterPro; IPR006594; Lish.			
DR	InterPro; IPR001680; WD40.			
DR	Pfam; PF00400; WD40; 8.			
DR	PRINTS; PR00320; GPROTEINBRPT.			
DR	ProDom; PD000018; WD40; 3.			
DR	SMART; SM00667; Lish; 1.			

34	294.5	10.8	643	1	T2D4_SCHPO	O13282 schizosacch
35	292.5	10.7	614	1	TU11_SCHPO	Q09715 schizosacch
36	291	10.7	395	1	YZLL_CAEEL	Q93847 caenorhabdi
37	289.5	10.6	395	1	GBB5_HUMAN	O14775 homo sapien
38	287.5	10.5	395	1	GBB5_MOUSE	P54314 mus musculus
39	285	10.5	586	1	TU12_SCHPO	Q9uug8 schizosacch
40	283	10.4	605	1	FW1A_HUMAN	O9v297 homo sapien
41	283	10.4	931	1	LUG_ARATH	O9fuy2 arabidopsais
42	281.5	10.3	554	1	PRP4_ARATH	O22212 arabidopsais
43	280.5	10.3	380	1	GBB_MAIZE	P49178 zea mays (m
44	280.5	10.3	713	1	TUPI_YEAST	P16649 saccharomyc
45	279	10.2	522	1	PRP4_HUMAN	O43172 homo sapien

DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00882; WD_REPEATS_2; 6.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT DOMAIN 4 36 LISH.
 FT REPEAT 167 206 WD 1.
 FT REPEAT 223 262 WD 2.
 FT REPEAT 264 303 WD 3.
 FT REPEAT 306 344 WD 4.
 FT REPEAT 347 386 WD 5.
 FT REPEAT 389 437 WD 6.
 FT REPEAT 440 479 WD 7.
 FT REPEAT 481 513 WD 8.
 FT DOMAIN 108 118 POLY-ALA.
 FT CONFLICT 31 31 E -> K (IN REF. 2).
 FT CONFLICT 59 59 Y -> H (IN REF. 3).
 FT CONFLICT 389 389 A -> Q (IN REF. 2).
 SQ SEQUENCE 514 AA; 55595 MW; 08556D2E4BA796D CRC64;
 Query Match 99.6%; Score 2716; DB 1; Length 514;
 Best Local Similarity 99.6%; Pred. No. 1.5e-188;
 Matches 512; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSISDEVNLFVRYLQBSGSHAFPGIKSHISQSNINGALVPPAALISITQKGLQYV 60
 DB 1 MSISDEVNLFVRYLQBSGSHAFPGIESHISQSNINGALVPPAALISITQKGLQYV 60
 QY 61 EAEVSNEDGTLFDRPIESLSLDVMPDVQVQYQYKDLAQQAQAAAAAASQ 120
 DB 61 EAEVSNEDGTLFDRPIESLSLDVMPDVQVQYQYKDLAQQAQAAAAAASQ 120
 QY 121 QGSAKNGENTANGENGAAHTIANNHDMEDVDGVEIPPNKAVLVRGHSEVFICANVPV 180
 DB 121 QGSAKNGENTANGENGAAHTIANNHDMEDVDGVEIPPNKAVLVRGHSEVFICANVPV 180
 QY 181 SDLLASGSGDSTARIWNLSENSTGSLVLRHICRGQDVPSNKDVTSLDWNSEGTLL 240
 DB 181 SDLLASGSGDSTARIWNLSENSTGSLVLRHICRGQDVPSNKDVTSLDWNSEGTLL 240
 QY 241 ATGSDYDGFARITWQGNLSTAGLQHKGFIFALKWKKGNFILSAGVDKTTIINDAHTGEA 300
 DB 241 ATGSDYDGFARITWQGNLSTAGLQHKGFIFALKWKKGNFILSAGVDKTTIINDAHTGEA 300
 QY 301 KOQPFHSAPALVDVQSNNTFASCTDMCHVCKLGDPRPIKTFQGHTEVNVNAIKWDPT 360
 DB 301 KOQPFHSAPALVDVQSNNTFASCTDMCHVCKLGDPRPIKTFQGHTEVNVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQNCVHDHQQHNKEIYTIKWSPTGPTGNPNANLMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSMKQNCVHDHQQHNKEIYTIKWSPTGPTGNPNANLMLASASF 420
 QY 421 DSTVRLWVDVDRGICHTLTKQBPVYVAFSPDGRYLASGDFKCVHWTQTGALVHSY 480
 DB 421 DSTVRLWVDVDRGICHTLTKQBPVYVAFSPDGRYLASGDFKCVHWTQTGALVHSY 480
 QY 481 RGTGGIFEVCVNNAAGDKVGASASGSCVCLDLRK 514
 DB 481 RGTGGIFEVCVNNAAGDKVGASASGSCVCLDLRK 514
 RESULT 2
 ID TLBX HUMAN STANDARD; PRG; 526 AA.
 AC O60907.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transducin beta-like 1X protein (Transducin-beta-like 1, X-linked).
 GN TLBX OR TBLX.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=99264241; PubMed=10330347;
 Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,
 Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.;
 "X-linked late-onset sensorineural deafness caused by a deletion
 involving OAI and a novel gene containing WD-40 repeats.",
 Am. J. Hum. Genet. 64:1604-1616(1999).
 [2]
 SEQUENCE FROM N.A.
 TISSUE=Lymph;
 MEDLINE=23388257; PubMed=12477932;
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettman M., Maman A., Rodriguez S., Sanchez A.,
 Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- SIMILARITY: Contains 1 Lish domain.
 CC -!- SIMILARITY: Contains 8 WD repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 EMBL; Y12781; CAA73319.1; ALT_INIT.
 EMBL; BC032708; AAH32708.1; -.
 Genew; HGNC:11585; TBLX.
 MIM; 300196; -.
 DR GO; GO:0007605; P:hearing; TAS.
 DR GO; GO:0007165; P:signal transduction; TAS.
 DR GO; GO:0007601; P:vision; TAS.
 DR GO; GO:0007601; P:vision; TAS.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001860; WD40.
 DR Pfam; PF04000; WD40; 8.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 7.
 DR PROSITE; PS50896; Lish; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00882; WD_REPEATS_2; 6.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT DOMAIN 4 36 LISH.
 FT REPEAT 179 218 WD 1.
 FT REPEAT 235 274 WD 2.
 FT REPEAT 276 315 WD 3.
 FT REPEAT 318 358 WD 4.
 FT REPEAT 359 398 WD 5.
 FT REPEAT 401 449 WD 6.
 FT REPEAT 452 491 WD 7.

```
FT REPEAT 493 525 WD 8.
FT DOMAIN 108 124 POLY-ALA.
SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F6E9 CRC64;

Query Match
Best Local Similarity 89.4%; Score 2437; DB 1; Length 526;
Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

QY 1 MSISDEVNFLVRYLQSGFHSFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSITSDVNFLVRYLQSGFHSFTFGIESHISQSNINGTLVPPAALISIIQKGLQYV 60
QY 61 EAEVINEEDGTLFDGRPIESLSIIDAVMPDVVQTRQAVRDKLAQOQAAAAA--- 117
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 EAEISINEEDGTLFDGRPIESLSIIDAVMPDVVQTRQAVRDKLAQOQAAAAA--- 117
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 -----ASQGSAXNGENTANGENGHAHTIANNHTDMVEVDGVEIIPNKAVLVRG 167
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 ATAATTAGVSHQNPKNREATVNGEENRAHSV-NNHAKPMEIDGVEIIPSKATVLRG 179
QY 168 HESVEFICAWNPNVSDLLASGSGDSTARIWNLSNSTSGTOLVLRHCIREGGQVPSNKG 227
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 HESVEFICAWNPNVSDLLASGSGDSTARIWNLSNSTSGTOLVLRHCIREGGHDPVPSNKG 239
QY 228 VTSLDWNSGTLATGSDYGFARITWDGNLSTLQGHKGFIFALKWNNKGNFILLSAGVD 287
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 VTSLDWNTGTLATGSDYGFARITWDGNLSTLQGHKGFIFALKWNNKGNFILLSAGVD 299
QY 288 KTTIWDHTGEAKQFPFHSAPALDVQSNNTFASCSTDMCHIVCKLGDRIKTFQG 347
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 KTTIWDHTGEAKQFPFHSAPALDVQSNNTFASCSTDMCHIVCKLGDRIKTFQG 359
QY 348 HTNEVNAIKWPTGNLLASCDDMTLKIWSKQDNCVHDLOHKNKEIYTIKWSPTGPGTN 407
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 HTNEVNAIKWPSGMLLASCDDMTLKIWSKQEVCHIDLQAHKEIYTIKWSPTGATS 419
QY 408 NPANMLASAFSTVRLWVDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 467
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 NPNSIMILASAFSTVRLWVDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 479
QY 468 IWTOTGALVHSYRGTTGGIFEVCAAGDKVGASASGSCVCLDLR 514
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 IWTQSGNLVHSYRGTTGGIFEVCAAGDKVGASASGSCVCLDLR 526

RESULT 3
ID TBLY_HUMAN
AC Q9BQ87.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Transducin beta-like 1Y protein (Transducin-beta-like 1, Y-linked).
GN TBLY OR TBLY.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,
RA Wilson R.K., Waterston R.H., Page D.C.;
RT "The DNA sequence of the human Y chromosome."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
```

or send an email to license@isb-sib.ch).

```
CC EMBL; AF332220; AAK13472.1; -
CC EMBL; AF332221; AAK13473.1; -
DR EMBL; AF332222; AAK13474.1; -
DR Genew; HGNC:18502; TBLY.
DR InterPro; IPR006594; Lish.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT DOMAIN 4 36 LISH.
FT REPEAT 177 216 WD 1.
FT REPEAT 233 272 WD 2.
FT REPEAT 274 313 WD 3.
FT REPEAT 316 354 WD 4.
FT REPEAT 357 396 WD 5.
FT REPEAT 399 447 WD 6.
FT REPEAT 450 489 WD 7.
FT REPEAT 491 521 WD 8.
SQ SEQUENCE 522 AA; 56688 MW; 4E020216422442D8 CRC64;

Query Match
Best Local Similarity 86.8%; Score 2367; DB 1; Length 522;
Matches 442; Conservative 35; Mismatches 34; Indels 12; Gaps 2;

QY 1 MSISDEVNFLVRYLQSGFHSFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSITSDVNFLVRYLQSGFHSFTFGIESHISQSNINGTLVPPAALISIIQKGLQYV 60
QY 61 EAEVINEEDGTLFDGRPIESLSIIDAVMPDVVQTRQAVRDKLAQOQAAAAA--- 115
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 EAEISINKGTVPDSPIESLSIIAVIPDVQMRQQAEGEKLQOQAAAAAASAMAX 120
QY 116 -----AASQGSAXNGENTANGENGHAHTIANNHTDMVEVDGVEIIPNKAVLVRGHE 169
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 AATMTAAISQQNPKNREATVNGEENGAHEI-NNHAKPMEIDGVEIIPNKATVLRGHE 179
QY 170 SEVFICAWNPVSDLLASGSGDSTARIWNLSNSTSGTOLVLRHCIREGGQVPSNKGDT 229
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 SEVFICAWNPVSDLLASGSGDSTARIWNLSNSTSGTOLVLRHCIREGGHDPVPSNKGDT 239
QY 230 SLDWNSGTLATGSDYGFARITWDGNLSTLQGHKGFIFALKWNNKGNFILLSAGVDKT 289
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 SLDWNSDGTLLAMGSDYGFARITWENGLASTLQGHKGFIFALKWNNKGNVILLSAGVDKT 299
QY 290 TIWDHTGEAKQFPFHSAPALDVQSNNTFASCSTDMCHIVCKLGDRIKTFQHT 349
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 TIWDHTGEAKQFPFHSAPALDVQSNNTFASCSTDMCHIVCKLGDRIKTFQHT 359
QY 350 NEVNAIKWPTGNLLASCDDMTLKIWSKQDNCVHDLOHKNKEIYTIKWSPTGPGTNP 409
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 NEVNAIKWPSGMLLASCDDMTLKIWSKQACVHDLDQAHKEIYTIKWSPTGATSNP 419
QY 410 NANMLASAFSTVRLWVDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 469
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 NSSIMILASAFSTVRLWVDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 479
QY 470 NTQTGALVHSYRGTTGGIFEVCAAGDKVGASASGSCVCLDL 512
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 NTQSGSLVHSYRGTTGGIFEVCAAGDKVGASASGSCVCLDL 522

RESULT 4
ID TBLY_MOUSE
PRT; 313 AA.
STANDARD;
```



```
DR PROSITE; PS50082; WD REPEATS 2; 4.
KW Repeat; WD repeat; Chromatin_regulator; Nuclear protein.
FT DOMAIN 4 36
FT REPEAT 163 208
FT REPEAT 218 257
FT REPEAT 259 298
FT REPEAT 315 354
FT REPEAT 357 396
FT REPEAT 399 437
FT REPEAT 440 479
FT REPEAT 496 535
FT CONFLICT 396 535
FT SEQUENCE 535 AA; 59145 MW; 25DD19AB2EFB4B07 CRC64;

Query Match 18.4%; Score 502.5; DB 1; Length 535;
Best Local Similarity 24.9%; Pred. No. 8.2e-29;
Matches 142; Conservative 100; Mismatches 232; Indels 97; Gaps 15;

QY 1 MSISDEVNFLVYRLOESGFSEA-----TFGIKSHISQSNINGALVPPAAL 49
DB 1 MSITSEELNLIWRYCOEMGHEVSALALQDETFRVLEFDEKYKEHI-----PLGTL 50
QY 50 ISITQKQGYVEAEVSEINEDG---TLFGRIEISLIDAVPDVWQTRQAVRDKLAQQ 106
DB 51 VNLVQRGILYTESELMVDYDSDIGALNEHHLSEDFNLVQALQID-----KEKPEI 101
QY 107 QAAAAAASQCGSAXNGENTANGENGHAHTIANNHDTMMEVDGVEIPPNKAVLVR 166
DB 102 SSEGRFTLETNSENKAGEDASTVERETQDDTNSIDSSD--DLIDGFVKI--LKEIV-- 155
QY 167 GHESEVFICAWPVSD--LLASGSDSTARINWLSNSTSGSTQ-----LVLRHCIREGG 219
DB 156 -KLDNIWSSWNPDESILAYGKNSVARLARIETDQEGKKYWKLIIEELRHPFALSA 214
QY 220 QDVPSNKDVTSLDWNSEGLTLATGSDGFARINTKDGKGLASTLQGHKGP;FALKWKKGN 279
DB 215 SSGKTTQVTCVLAWSHDGNSIVTVGNGELRLNKTGALLNLVNFHRAPIVSVKWKDGT 274
QY 280 PILSAGVDKTTIWDANTGEAKQOFPF-----HSAP---ALDWDQSNNTFAS 324
DB 275 HIISMDVENVTILMNVI-SGTVMQHFELKETGGSSINAEHSGDGLVDVVEWDDDKFVI 334
QY 325 CSTDMCHVCKLQGDRIKTPQGTNEVNAIKWPTGNILLASCDDMTLKWSMKQDNCV 384
DB 335 PGPKGALFVQITEKTTGKLIGHHGPISVLEFNDTKLLLSASDDGTLLRWHGGNSQ 394
QY 385 HDLQHNKEIYTIKWSPTGPTNPNANMLASASFDSTVRLWVDVDRGICIHILTKHQEP 444
DB 395 NSFYGHSSQIVSASWV-----GDDKVISSCMDGVSRLWSLKQNTLLALSIVDGV 444
QY 445 VYSVAFSPDGRYLASGDFKCVHIW-----NTGTAL-----VHSYR 481
DB 445 IFAGRIQDQGYKAYAFMDGVGVVYDLKLNKRSRLYGNRDGLINLPDIPLYASYOSSQ 504
QY 482 GTGGEIFVCVNAAGDKVGA--SASDGSVCVL 510
DB 505 DNDVIFDLNWCAGNKISVAYSLOEGSVVAI 535

RESULT 6
YY46_ANASP
ID YY46_ANASP STANDARD; PR7; 1526 AA.
AC QBYRIL;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein alr3466.
GN ALR3466.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RL DNA Res. 8:205-213(2001).
CC -1- SIMILARITY: Contains 16 WD repeats.
CC -1- SIMILARITY: Contains 1 pentapeptide repeat domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF003593; BAB75165.1; -.
DR PIR; AC2239; AC2239.
DR InterPro; IPR007111; NACT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 15.
DR PRINTS; PRO0320; GPROTEINRET.
DR PRODOM; PD000018; WD40; 8.
DR SMART; SM00320; WD40; 15.
DR PROSITE; PS00678; WD REPEATS 1; 11.
DR PROSITE; PS00082; WD REPEATS 2; 15.
DR PROSITE; PS02394; WD REPEATS REGION; 1.
DR PROSITE; PS0837; NACT; UNKNOWN 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 334 376
FT DOMAIN 823 862 PENTAPEPTIDE.
FT REPEAT 862 901 WD 2.
FT REPEAT 904 945 WD 3.
FT REPEAT 945 985 WD 4.
FT REPEAT 985 1027 WD 5.
FT REPEAT 1030 1069 WD 6.
FT REPEAT 1072 1111 WD 7.
FT REPEAT 1114 1153 WD 8.
FT REPEAT 1156 1195 WD 9.
FT REPEAT 1198 1237 WD 10.
FT REPEAT 1240 1279 WD 11.
FT REPEAT 1282 1321 WD 12.
FT REPEAT 1324 1363 WD 13.
FT REPEAT 1366 1405 WD 14.
FT REPEAT 1408 1447 WD 15.
FT REPEAT 1450 1491 WD 16.
SQ SEQUENCE 1526 AA; 170376 MW; D5DF272509B4A738 CRC64;

Query Match 17.6%; Score 479; DB 1; Length 1526;
Best Local Similarity 29.8%; Pred. No. 1.6e-26;
Matches 117; Conservative 76; Mismatches 149; Indels 50; Gaps 10;

QY 136 NGAHTIANNHNTDMEVDGVEIPPNKAV-VLRGHESEVFICAWNPNVSDLLASGSGDSTAR 194
DB 1127 NGV-TLNGSSQIVRLWD--ISSKKCLYTLQHTNWNNAVAFSPDGTALASGSDQTVR 1183
QY 195 IWNLSNS-----TSGSTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGLTLATGSDGFAR 250
DB 1184 LWDISSKKCLYILQHTSW-----VNSVVFENPDGSLASGSDQTVR 1225
QY 251 IW-TRDGNLSTLQGHKGP;FALKWKKGNFTLSAGVDKTTIWDHAHTGEAKQQPFHSA 309
DB 1226 LWEINSSKCLCTFQGHTSWNSVFNPDGSMLASGSDKTVRLWDISSKCLHTFQGH- 1284
QY 310 PALDWDQSNNTF-----ASCSTDMCHVCKLQGDRIKTPQGTNEVNAIKWPTGN 362
DB 1285 -----NWNVSVAFNPDGSMLASGSDQTVRLWEISSKCLHTFQGHTSWNSVTFSPDGT 1339
QY 363 LLASCSDDMTLKWSMKQDNCVHDLQHNKEIYTIKWSPTGPTNPNANMLASASFD 422
```

DB 1340 MLAGSDQTVRLWSISSGECGLYFLGHTNNGVIFSPDG-----AILASGSDQ 1390
QY 423 TVRLWDVDRGICHTLTKHQFPVYAFSPDGRYLAGSPDKCVHINWNTOTGALVHSYRG 482
DB 1391 TVRLWSISSGCLYTLQGNHNNWGSIVFSPDGTLLASGSDQTVRLWNISGECGLYTLHG 1450
QY 483 -TGGIFFCVWNAAGKVCASASDGSVCVLDLR 513
DB 1451 HINSVRSVAFSSDGLILASGSDDETIKLWDVK 1482
RESULT 7
ID YS00 ANASP STANDARD; PRT; 1258 AA.
AC Q8YTC2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical WD-repeat protein alr2800.
GN ALR2800.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- SIMILARITY: Contains 15 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF003590; BAB74499.1;
CC PIR; A12155; A12155.
CC InterPro; IPR002182; NB-ARC.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00931; NB-ARC; 1.
CC Pfam; PF00400; WD40; 14.
CC PRINTS; PR00320; GPROTEINBRPT.
CC ProDom; PD000018; WD40; 1.
CC SMART; SM00320; WD40; 14.
CC PROSITE; PS00678; WD_REPEATS_1; 9.
CC PROSITE; PS00682; WD_REPEATS_2; 14.
CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 55 93
FT REPEAT 640 679
FT REPEAT 692 721
FT REPEAT 724 763
FT REPEAT 766 807
FT REPEAT 809 849
FT REPEAT 850 889
FT REPEAT 892 931
FT REPEAT 934 975
FT REPEAT 976 1017
FT REPEAT 1019 1059
FT REPEAT 1060 1101
FT REPEAT 1103 1143
FT REPEAT 1144 1183

FT REPEAT 1186 1227 WD 15.
SQ SEQUENCE 1258 AA; 139513 MW; 45DF03B91170C451 CRC64;
Query Match 17.0%; Score 463.5; DB 1; Length 1258;
Best Local Similarity 31.8%; Pred. No. 1.7e-25;
Matches 103; Conservative 44; Mismatches 100; Indels 77; Gaps 7;
QY 165 LRHSEVEFICAWNPSVLLASGSDSTARIWNLSGNSSTGSLVLRHRCIREGQDVPS 224
DB 974 LEGHTDFYIGIAFSPDSQTLASASTDSSVRLWNI---STGQCFQLLEH----- 1019
QY 225 NKDVTSLDW-----NSEGTLATSGYDGFARIWT-KDGNLASTLGQKHQPIFALKWNKK 277
DB 1020 -----TDWYAVVPHPOQGIATGSADCTVKLWNIISGQCLTKLSEHSDKILGNWSPD 1073
QY 278 GNFIISAGVDKTTIWDATHTGEAKQFPHPSPALPDVDWQSNNTFASCTDMCIHVCKLG 337
DB 1074 GQLLASASADQSVRLWD-----CCTGRCVGI----- 1099
QY 338 QDRPIKTFOGHTNEVNAIKWDPTGNLLASCDSDMTLKIWSMKQDNCVHDLQOHNKEIYTI 397
DB 1100 -----LRGHSNRVYSAIFSPNGEIIATCSDTQVKIWDWQQKCLTKLTGHTNWVFDI 1152
QY 398 KWSPTGPTGNNPNANMLASAFDSFVRLWDYDRGICHTLTGKHQEPYVSVAFSFGRYL 457
DB 1153 AFSPDGK-----ILASASHDQTVRIWDVNTGKCHHICITGHTLVSSVAFSPDGEV 1203
QY 458 ASGSPDKCVHINWNTOTGALVHSYR 481
DB 1204 ASGSDQTVRIWNVKTGECQLILR 1227
RESULT 8
HET1 PODAN
ID HET1 PODAN STANDARD; PRT; 1356 AA.
AC Q0808;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vegetative incompatibility protein HET-E-1.
GN HET-E1.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=96009891; PubMed=7557402;
RA Saube S., Turcq B., Begueret J.;
RT "A gene responsible for vegetative incompatibility in the fungus
RT Podospora anserina encodes a protein with a GTP-binding motif and G
RT beta homologous domain.";
RL Gene 162:135-139(1995).
CC -!- FUNCTION: RESPONSIBLE FOR VEGETATIVE INCOMPATIBILITY THROUGH
CC SPECIFIC INTERACTIONS WITH DIFFERENT ALLELES OF THE UNLINKED GENE,
CC HET-C.
CC -!- SIMILARITY: Contains 10 WD repeats.
CC -!- SIMILARITY: Contains 1 NACHT domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L28125; AAA85775.1;
CC PIR; T18521; T18521.
CC InterPro; IPR007111; NACHT_NTPase.
CC InterPro; IPR001680; WD40.
CC Pfam; PF05729; NACHT; 1.
CC Pfam; PF00400; WD40; 10.
DR
DR
DR
DR
DR


```

DR PRINTS; PRO0320; GPROTEINBRPT.
DR PRODOM; PD000018; WD40; 10.
DR SMART; SMO0320; WD40; 10.
DR PROSITE; PS00837; NACHT; 1.
DR PROSITE; PS00678; WD REPEATS 1; 10.
DR PROSITE; PS00082; WD REPEATS 2; 10.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW GTP-binding; Repeat; WD repeat.
FT DOMAIN 294 629 NACHT.
FT NP_BIND 300 307 GTP (POTENTIAL).
FT REPEAT 839 869 WD 1.
FT REPEAT 891 911 WD 2.
FT REPEAT 923 953 WD 3.
FT REPEAT 965 995 WD 4.
FT REPEAT 1007 1037 WD 5.
FT REPEAT 1049 1079 WD 6.
FT REPEAT 1091 1121 WD 7.
FT REPEAT 1133 1163 WD 8.
FT REPEAT 1175 1205 WD 9.
FT REPEAT 1217 1247 WD 10.
SQ SEQUENCE 1356 AA; 149765 MW; 965FB319844E0651 CRC64;

Query Match 16.9%; Score 460; DB 1; Length 1356;
Best Local Similarity 27.9%; Pred. No. 3.3e-25;
Matches 124; Conservative 73; Mismatches 178; Indels 70; Gaps 12;

QY 78 IESLSLIDVMPDVVQTRQAYRDKLAQQQAAAAAASQGSAGNKE-----NT 130
DB 820 ISTSVVEAEWNACTQT-----LEHGSSVLSVAFADGQVAGSGDDTKTIKWDT 870
QY 131 ANGEENGAAHTIANHNTDMVEVDGVEIPPNAKAVLRGHESEVFICAWNPVSDLLASGSD 190
DB 871 ASG--TGTQT-----LEHGGSVMSVAFSPDRVRVAGSD 904
QY 191 STASINWLSNSTSGS--TQVLRHCIREGGQDVPSNKDVTSLDWNSEGTLLATGSDGPA 249
DB 905 KTIKIW-----DAASGTCQTGLEH-----GGR-----VQSVAFSPDQGVAGSGDDHTI 949
QY 250 RIW--TKDGNLASTLGQHKGPFFALKWNKGNFILSAGVDKTTIWDHAHTGEAKQGFPHS 308
DB 950 KIWDAASTCTQTGLEHGSVLSVAFSPDQGVAGSGDKTIKIWDTSCTCTQTTLEHG 1009
QY 309 APALDVNQ--SNNTFASCTDMCHVCKLGDORPIKTFQHTNEVNAIKWDPGNLIASC 367
DB 1010 GSVMSVAFSPDQGVAGSGDDTKIKIWDTSCTCTQTTLEHGGVWQSVFSPDQGVASG 1069
QY 368 SDDMTLKISWKQDNCVHDLQHNKEIYTIKWSPTGPTNPNANLMLASASFDSTVRLW 427
DB 1070 SDDHTIKIWDVSGTCTQTTLEHGSVMSVAFSPDQ-----RVAGSIDGTIKIW 1120
QY 428 DVDRGICHTLTKHQEPYISVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSYRGTTG-I 486
DB 1121 DAASGTCQTQTTLEHGGVWHSVAFSPDQGVAGSGIDGTIKIWDAASTCTCTQTTLEHGGW 1180
QY 487 FEVCWNAAGDKVGASADSGVCVLD 511
DB 1181 QSVAFSPDQGVAGSGDDTKIKIWD 1205

RESULT 9
YL24 ANASP
ID YL24 ANASP STANDARD; PRT; 1683 AA.
AC Q8V57;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein al12124.
GN AL12124
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: Contains 13 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003588; BAB73823.1; -
DR F1R; AF2071; AF2071. -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 14.
DR PRINTS; PRO0320; GPROTEINBRPT.
DR SMART; SMO0320; WD40; 14.
DR PROSITE; PS00678; WD REPEATS 1; 3.
DR PROSITE; PS00082; WD REPEATS 2; 14.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1070 1109 WD 1.
FT REPEAT 1111 1150 WD 2.
FT REPEAT 1153 1192 WD 3.
FT REPEAT 1195 1236 WD 4.
FT REPEAT 1280 1318 WD 5.
FT REPEAT 1320 1359 WD 6.
FT REPEAT 1361 1400 WD 7.
FT REPEAT 1402 1441 WD 8.
FT REPEAT 1444 1483 WD 9.
FT REPEAT 1486 1525 WD 10.
FT REPEAT 1528 1567 WD 11.
FT REPEAT 1570 1609 WD 12.
FT REPEAT 1611 1650 WD 13.
SQ SEQUENCE 1683 AA; 185459 MW; F7CB361BF54F7137 CRC64;

Query Match 16.3%; Score 445.5; DB 1; Length 1683;
Best Local Similarity 28.5%; Pred. No. 4.9e-24;
Matches 109; Conservative 69; Mismatches 165; Indels 40; Gaps 6;

QY 124 AKNGENTANGENGAAHTIANHNTDMVEVDGVEIPPNAKAVLRGHESEVFICAWNPVSDL 183
DB 1290 SSDGKAIASASADNTIKLWNRHGLE-----TFTGSGGVYAVNFDPDSNI 1336
QY 184 LASGSGDSTARIWLSNSTSGSTQVLRHCIREGGQDVPSNKDVTSLDWNSEGTLLATG 243
DB 1337 IASASLDNTIRLWQRPISPL-----EVLAGNSGVYAVSFLHDGSIATA 1381
QY 244 SYDGFARIW--TKDGNLASTLGQHKGPFFALKWNKGNFILSAGVDKTTIWDHAHTGEAKQ 302
DB 1382 GADGNTQLWHSQDGSLLKTLPGNKA--IYGISFTPGQDLIASANADKTKIWRVDRGKALK 1440
QY 303 QPFHSAPALDVNQ--SNNTFASCTDMCHVCKLGDORPIKTFQHTNEVNAIKWDPG 361
DB 1441 TLIGHDNEVKNVNFSPDGKTLASASRDNTVKNWVSDGKFKKTLKGHTDEVFWVFSFPG 1500
QY 362 NLLASCSDDMTLKISWKQDNCVHDLQHNKEIYTIKWSPTGPTNPNANLMLASASFD 421
DB 1501 KIIASASADKTIKRLWDSFSGNLKISLPAHNDLVYSVNFNPDGS-----MLASTSAD 1551
QY 422 STVRLWVDVRGICHTLTKHQEPYISVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSYR 481
DB 1552 KTVKLWRSHDGHLLHTFSGHSNNVYSSSPDGRVIASASEDKTKIWIQDGHLLTLPQ 1611
QY 482 GTGGIFEVCWNAAGDKVGASASD 504

```

```

Db 1612 HQAGVMSAIFSPDGRITLISGLD 1634
RESULT 10
Y163_SYN3
ID Y163_SYN3 STANDARD; PRT; 1693 AA.
AC Q55563;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical WD-repeat protein sil0163.
GN SL0163.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -!- SIMILARITY: Contains 16 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D63999; BAA10064.1; -
CC PIR; S76086; S76086.
CC InterPro; IPR008941; TPR-like.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 16.
CC PRINTS; PR00320; GPROTEINRPT.
CC ProDom; PD000018; WD40; 13.
CC SMART; SM00320; WD40; 16.
CC PROSITE; PS00678; WD REPEATS_2; 15.
CC PROSITE; PS00082; WD REPEATS_2; 15.
CC PROSITE; PS02094; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1008 1042 WD 1.
FT REPEAT 1053 1083 WD 2.
FT REPEAT 1094 1124 WD 3.
FT REPEAT 1135 1165 WD 4.
FT REPEAT 1176 1206 WD 5.
FT REPEAT 1217 1247 WD 6.
FT REPEAT 1258 1288 WD 7.
FT REPEAT 1299 1329 WD 8.
FT REPEAT 1340 1370 WD 9.
FT REPEAT 1381 1411 WD 10.
FT REPEAT 1422 1452 WD 11.
FT REPEAT 1463 1493 WD 12.
FT REPEAT 1504 1534 WD 13.
FT REPEAT 1545 1575 WD 14.
FT REPEAT 1586 1616 WD 15.
FT REPEAT 1627 1657 WD 16.
SQ SEQUENCE 1693 AA; 09778827A0251CFF CRC64;
Query Match
Best Local Similarity 14.3%; Score 388.5; DB 1; Length 1693;
Matches 107; Conservative 52; Mismatches 138; Indels 79; Gaps 10;
QY 100 RDKLAQQQAAAAAASQGGAKNGCENTANGE--ENGHTTANNHTDMVEVDGVEI 157
DB 1362 RDKTARLTTEGECVAVLADHGWRGQFSPDQGVITGS--ADKTAQLNVLG---- 1414

```

RESULT 11

```

HUS7_HUMAN
ID HUS7_HUMAN STANDARD; PRT; 485 AA.
AC Q9NVX2; O60868; Q9BU54;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE WD-repeat protein HUSSY-07.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tsogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Magatsuna M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RA "NEDO human cDNA sequencing project";
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung, and Placenta;
RC MEDLINE=22388257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RX Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RX Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RX Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh F.,
RX Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RX Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RX Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RX Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RX Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RX Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RX Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RX Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RX Whiting M., Madan A., Young A.W., Green E.D., Bouffard G.G.,
RX Blakesley R.W., Touchman J.W., Shevchenko Y., Bouffard G.G.,
RX Rodriguez A.C., Grimwood J., Schmitz J., Myers R.M.,
RX Butterfield Y.S.N., Krzywinski M.I., Skalska J., Smaluk D.E.,
RX Schnzerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RN [3]

```

SEQUENCE OF 400-485 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21064499; PubMed=11124703;
RA Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,
RA Canata N., Zimbello R., Lanfranchi G., Valle G.;
RT "Characterization of 16 novel human genes showing high similarity to
RT yeast sequences.";
RL Yeast 18:69-80(2001).
RN [4]
RP SUBCELLULAR LOCATION
RX MEDLINE=22317277; PubMed=12429849;
RA Scherl A., Coute Y., Deon C., Calle A., Kindbeiter K., Sanchez J.-C.,
RA Greco A., Hochstrasser D.F., Diaz J.-J.;
RT "Functional proteomic analysis of human nucleolus.";
RL Mol. Biol. Cell 13:4100-4109(2002).
CC -|- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -|- SIMILARITY: Contains 8 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AK001320; BAA91621.1; -
DR EMBL; BC002884; AAHC02884.2; -
DR EMBL; BC012075; AAH12075.1; -
DR EMBL; AJ005257; CAA06444.1; -
DR SWISS-2DPAGE; OSNVX2; HUMAN.
DR InterPro; IPR001632; Gprotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PROSITE; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS50082; WD REPEATS 2; 7.
DR PROSITE; PS50294; WD REPEATS REGION; 1.
KW Nuclear protein; Repeat; WD repeat.
FT REPEAT 112 151
FT REPEAT 154 193
FT REPEAT 197 241
FT REPEAT 244 282
FT REPEAT 325 366
FT REPEAT 370 409
FT REPEAT 412 451
FT REPEAT 454 484
SQ SEQUENCE 485 AA; 53266 MW; 0A8BE24B4957379 CRC64;
Query Match 13.4%; Score 366.5; DB 1; Length 485;
Best Local Similarity 26.8%; Pred. No. 4.7e-19;
Matches 103; Conservative 47; Mismatches 145; Indels 89; Gaps 9;
QY 165 LRGHESVFICANVPVSDLLASGSGDSTARINWSEN----STFGSTQLVLRHCIREGGQ 220
DB 110 LEHSEAVISVAFSPGKYLASGGDTIVFWDLSHTPTFTCKGRHWLVSWSVDPGK 169
QY 221 DVPS-----NKDVTSLDW-----NSEGTLATSYDGFAR 250
DB 170 KLASGCKNGQILLWDPSTGKQVGRITAGSKWITGLISWEPLHANPECRYVAVSSKDGVR 229
QY 251 IW-TKDGNLASTLQHGKPIFALKWKKGNFILSAGVDKTIIDAHTEAKQFPFH-- 307
DB 230 IWDTTAGRCRILTGHTQSVTCURWGGDG-LLYSASQDRTIKVRAHDGVLCTLQGHGH 288
QY 308 --SAPALVDVW-----QSNNTFASCS 326
DB 289 WYNTMALSTDYALRTGAFEPASVNPQDLQGSQELKERALGRYINLVRRGQGFRLVSGS 348
QY 327 TDMCIHVCKLQGD-RPIKTFQGHTEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQDNCVH 385

DB 349 DQFTFLFMSPAEDKKKPLRTMTGHOALINQVLFSPDSRIVASASFDKSIKLDGRTGKYLA 408
QY 386 DLQQHNKEIYIKWSPGTGFTNNPNANMLASASFDSTVRLWDVDRGICHTLTKHQEPV 445
DB 409 SURGHVAAVYQIANSAD-----SRLVSGSDSTLKWVDVRAQKLAAMDLPQHADEV 459
QY 446 YSVAFSPDGRVSLASGDFKCVHIW 469
DB 460 YAVDWSPDQRVASGGKDKLRW 483
RESULT 12
HUS7 MOUSE
ID HUS7 MOUSE STANDARD; PRT; 485 AA.
AC Q8VET4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE WD-repeat protein HUSY-07.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguercio N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.W., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting R.W., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -|- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -|- SIMILARITY: Contains 8 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; BC018399; AAH18399.1; -
DR InterPro; IPR001632; Gprotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS50082; WD REPEATS 2; 7.
DR PROSITE; PS50294; WD REPEATS REGION; 1.
KW Nuclear protein; Repeat; WD repeat.
FT REPEAT 112 151
FT REPEAT 154 193
FT REPEAT 197 241
FT REPEAT 244 282

FT REPEAT 197 241 WD 3.
 FT REPEAT 244 282 WD 4.
 FT REPEAT 327 366 WD 5.
 FT REPEAT 370 409 WD 6.
 FT REPEAT 412 451 WD 7.
 FT REPEAT 454 484 WD 8.
 SQ SEQUENCE 485 AA; 53148 MW; 207595D603AEC07B CRC64;

Query Match 13.4%; Score 365.5; DB 1; Length 485;
 Best Local Similarity 26.6%; Pred. No. 5.5e-19;
 Matches 102; Conservative 48; Mismatches 145; Indels 89; Gaps 9;

QY 165 LRHSESEYFICAMPVPSDLASGSGDSTARIWNLSN-----STSGTQLVLRHCIRGGQ 220
 DB 110 LEHSEAVISAFSPTGKYLAGSGDITVRFWDLSTETPHTCKGHRHWLSISWSPDGK 169
 QY 221 DVPS-----NKDVTSLDW-----NSEGTLATGSYDGFAR 250
 DB 170 KLASCKNGQILLWDPSTGLQVGRTLTGHSKWITGLSWEPLHMNPECRYVASSKDGSVR 229
 QY 251 IW-TKGNLSTLQGHKGPPIFALKWKNKGNFILSAGVDKTTIWDATHTGEAKQOPPH-- 307
 DB 230 VMDTTAGCEBILTGHTOSVTLRAGDG-LLYSASQDRITKWRANDGVLCRTLQGHG 288
 QY 308 --SAPALDVM-----QSNNTFASCS 326
 DB 289 WNTWTALSTDVALRTGAPAEATVNAQDLOGSLKELKERASSRYNLVRGQPERLVSGS 348
 QY 327 TDMCIHVCKLQD-RPIKTFQGHNEVNAIKWDPGTNLLASCSDMTLKIWSKQDNCVH 385
 DB 349 DDFTLFLMSPAEDKPLARMTGHOALINQVLFSPDSRIVASASPKSIKLDWGTGKYLA 408
 QY 386 DLQCHKEIYIKWSPGTGNTNPNANMLASASFDSTVRLWDVDRGICHTITKHOEPV 445
 DB 409 SLRGHVAAYVQLASAD-----SRLVSGSSDSTLKWVDKAKLATLDPGHADV 459
 QY 446 YSVAFSPDGRVLASGSPKCVHIW 469
 DB 460 YAVDMSPDQQRVASGGKDKCLRIV 483

RESULT 13

YCW2_YEAST STANDARD; PRT; 515 AA.
 AC P25382; Q8NKG4;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical WD-repeat protein YCR072C.
 GN YCR072C OR YCR72C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=92244356; PubMed=1574125;
 RA Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
 RA Alverghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
 RA Benoit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
 RA Botin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
 RA Carginani G., Chanet R., Contreras R., Crouzet M., Daigian-Fornier B.,
 RA De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
 RA Dujon B., Duesterhoft A., Erdmann D., Esteban M., Fiers W.,
 RA Fairhead C.A., Faye G., Feldmann H., Fiers W.,
 RA Francinguet-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
 RA Fuller L.J., Gent M.E., Gigot D., Galliquet V., Glansdorff N.,
 RA Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haesemann M.,
 RA Hatat D., Hegemann J.H., Herbert C.J., Hilger F., Hohmann S.,
 RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
 RA Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
 RA Kleinhaus U., Kreisl P., Lafranchi G., Lewis C., van der Linden C.G.,

RA Lucchini G., Lutzenkirchen K., Maat C., Mannhaupt G., Manzano M.E.,
 RA Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
 RA Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
 RA Newton C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
 RA Pereira J., Philippsen P., Pierard A., Planta R.J., Plevani P.,
 RA Poetsch B., Pohl P.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
 RA Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
 RA Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
 RA Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
 RA Spiegelberg R., Strateva L.I., Steensma H.Y., Steiner S., Thierry A.,
 RA Thirion G., Triano L.N., Urrutiarazu L.A., Vallie G., Vetter I.,
 RA van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warmington J.R.,
 RA von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
 RA Zimmermann F.K., Sgourou J.G.;
 RT "The complete DNA sequence of yeast chromosome III.";
 RL Nature 357:38-46(1992).
 RN [2]
 RP REVISIONS.
 RA Jimenez A.;
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP REVISION TO 444.
 RA Valles G., Volckaerts G.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Contains 8 WD repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X59720; CAC42989.1; -.
 DR PIR; S19487; S19487;
 DR GERMOnline; 138973; -.
 DR SGD; S0000668; YCR072C.
 DR InterPro; IPR001632; Gpoteins_B.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF04000; WD40; 8.
 DR PRINTS; PR00319; GPROTEINB.
 DR PRINTS; PR00320; GPROTEINBPT.
 DR PRODom; PD000018; WD40; 6.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00678; WD_REPEATS_1; 5.
 DR PROSITE; PS00682; WD_REPEATS_2; 7.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Hypothetical protein; Repeat; WD repeat.
 FT REPEAT 141 181
 FT REPEAT 184 223
 FT REPEAT 227 273
 FT REPEAT 276 314
 FT REPEAT 352 396
 FT REPEAT 400 439
 FT REPEAT 442 481
 FT REPEAT 484 514
 FT REPEAT 515 514
 SQ SEQUENCE 515 AA; 57025 MW; 04451B0C38FB5318 CRC64;

Query Match 13.4%; Score 365.5; DB 1; Length 515;
 Best Local Similarity 23.2%; Pred. No. 6e-19;
 Matches 124; Conservative 85; Mismatches 164; Indels 161; Gaps 22;

QY 38 NINGALVPPAALISITOKGLQYVEAEVSNEDGT-----FDGR-----PIESLSLIDA 86
 DB 44 NYGGLVPPGA---ISEKLELLNLQNGTSDPPYTFCTIOGKASDPVKITIDTN 100
 QY 87 VNPDVVTRQAYRKLQAQQAQAAAAAASQGSANGKNGTANGENGAIHANNHT 146
 DB 101 LYSSLIKPGYNSTEDQITLLYTPRAVFKVPVTRSSA----- 138
 QY 147 DNMEVDGDVEIPPNKAVLVRGHESEVFICAMNP-VSDLLASGSGDSTARIWNLSN-----N 201


```

RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Marra M.A.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=99367400; PubMed=10438340;
RA Johnstone R.W., Tommerup N., Hansen C., Vissing H., Shi Y.;
RT "Structural organization, tissue expression, and chromosomal
RT localization of C10A1, a functional modulator of the Wilms' tumor
RT suppressor, WT1.";
RL Immunogenetics 49:900-905(1999).
CC -!- FUNCTION: Seems to specifically modulate the transactivation
CC activity of WT1.
CC -!- SUBUNIT: Interacts with WT1.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: 'C10A' means 'bridge' in Chinese.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U63810; AAC24948.1; -;
DR EMBL; AC004020; AAC23493.1; -;
DR EMBL; EC001395; AAH01395.1; -;
DR EMBL; EC032812; AAH32812.1; -;
DR MIM; 604333; -;
DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
DR GO; GO:0006357; P:regulation of transcription from Pol II promoter; TAS.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40.7
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 2.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat; Nuclear protein.
FT REPEAT 14 53 WD 1.
FT REPEAT 59 98 WD 2.
FT REPEAT 103 142 WD 3.
FT REPEAT 148 187 WD 4.
FT REPEAT 192 231 WD 5.
FT REPEAT 250 289 WD 6.
FT REPEAT 301 338 WD 7.
SQ SEQUENCE 339 AA; 37840 MW; 63A8D8257A204FC8 CRC64;

Query Match 12.8%; Score 348; DB 1; Length 339;
Best Local Similarity 30.4%; Pred. No. 6.3e-18;
Matches 92; Conservative 42; Mismatches 89; Indels 80; Gaps 11;

```

```

QY 222 VPSNKDVTS--LDWNSBGLLATGYDGFARITWKDGN--LASTLGO-HKGPIFALKWN 275
DB 11 VFAHPDSRCWFLAWNPAAGTLLASCGRRIIRIWGTGDSWICKSVLSBHGORTVRKVAWS 70
QY 276 KKGNTLSAGVDKTTIIWDAHTGEAKQOPPHSAFALDQVDSNNTPASCSTDMCIHVCK 335
DB 71 PCGNVLASASPDATTCI-----WKQDQDFECVT----- 99
QY 336 LQQDRPIKTPQGHTEVNAIKWDPTGNLLASCSDDMTLKIMSKODN--CVHDLQOHNK 392
DB 100 -----TLECHENEVKSVAWAPSGNLLATCSRDKSVWVWEVDEDEYECVSLNSHTQ 151
QY 393 EYITIKWSPGTGNTNPNANMLASASFDSTVRLW--DVDRGICHTLTKEOEPVYSVAF 450
DB 152 DVKHVVVWHP-----SQELLASASYDDTVKLYREBEDDWVCCATLEGHSTVWSLAF 202
QY 451 SPDGRYLASGSPDKCKVHIWNTOTGALVHSYRGTTGGIFEVCVWNAAGDKYVGASADGS---V 507
DB 203 DFGQRLASCSDRTVRIWRYLPG-----NEQG--VACSGSDPSWKCI 244
QY 508 CVL 510
DB 245 CTL 247

Search completed: August 9, 2004, 16:46:10
Job time : 18.6667 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:40:42 ; Search time 8.66667 Seconds
(without alignments)
3088.161 Million cell updates/sec

Title: US-09-987-701-12
Perfect score: 2731
Sequence: 1 MSISDEVFLVRYLQSSG.....GDKVGASASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2696	98.7	514	1 TELR_HUMAN	Q9bzk7 homo sapien
2	2439	89.3	526	1 TELX_HUMAN	O60907 homo sapien
3	2351	86.1	522	1 TELY_HUMAN	Q9bq87 homo sapien
4	1304	47.7	313	1 TELX_MOUSE	Q9qxe7 mus musculus
5	501.5	18.4	535	1 SIF2_YEAST	P38262 saccharomyc
6	481	17.6	1526	1 VY46_ANASP	Q8yrr1 anabaena sp
7	462	16.9	1356	1 HET1_PODAN	Q00808 podospora a
8	439.5	16.8	1258	1 Y500_ANASP	Q8ytc2 anabaena sp
9	451	16.5	1683	1 YL24_ANASP	Q8yvt7 anabaena sp
10	390.5	14.3	1693	1 Y163_SYNY3	Q55563 synchocyst
11	375.5	13.7	515	1 YCW2_YEAST	P25382 saccharomyc
12	363.5	13.3	485	1 HUS7_HUMAN	Q9nvx2 homo sapien
13	362.5	13.3	485	1 HUS7_MOUSE	Q8vej4 mus musculus
14	358	13.1	437	1 A3C3_DICDI	P14197 dictyosteli
15	349	12.8	339	1 CIAL1_HUMAN	O76071 homo sapien
16	338.5	12.4	742	1 PKWA_THRCU	P49695 thermomonos
17	336.5	12.3	409	1 LIS1_HUMAN	P43034 homo sapien
18	335.5	12.3	409	1 LIS1_BOVIN	P43033 bos taurus
19	335.5	12.3	800	1 T2D4_HUMAN	Q15542 homo sapien
20	334.5	12.2	409	1 LIS1_MOUSE	P43035 mus musculus
21	331.5	12.1	334	1 WDS5_HUMAN	Q9g9p9 homo sapien
22	328.5	12.0	361	1 MDR5_DROME	Q9v378 drosophila
23	328.5	12.0	606	1 PF20_CHLRE	P93107 chlamydomon
24	326	11.9	798	1 T2D4_YEAST	P38129 saccharomyc
25	324.5	11.9	579	1 SE10_CABEL	Q93794 caenorhabdi
26	313	11.5	640	1 MT30_YEAST	P39014 saccharomyc
27	312	11.4	376	1 YKY4_CABEL	Q17963 caenorhabdi
28	310	11.4	1191	1 Y143_SYNY3	P74442 synchocyst
29	307	11.2	643	1 T2D4_SCHPO	O13282 schizosacch
30	306.5	11.2	704	1 T2D4_DROME	P49846 drosophila
31	306	11.2	678	1 SCOB_EMENI	Q00659 emericella
32	303.5	11.1	501	1 YH92_CABEL	Q23256 caenorhabdi
33	302	11.1	605	1 POF1_SCHPO	P87053 schizosacch

ALIGNMENTS

RESULT 1

ID	TELR_HUMAN	STANDARD;	PRT;	514 AA.
AC	Q9BZK7; Q9H2I1; Q9H9A1;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Nuclear receptor co-repressor/HDAC3 complex subunit TBLR1 (TBLR1-related protein 1).			
GN	TBLR1 OR IRA1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND INTERACTION WITH HDAC3.			
RX	MEDLINE=21930350; PubMed=11931768;			
RA	Zhang J., Kalkum M., Chait B.T., Roeder R.G.;			
RT	"The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2.";			
RL	Mol. Cell 9:1611-623 (2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Zhang X., Dormady S., Basch R.;			
RT	"Identification of four human cDNAs that are differentially expressed by early hematopoietic progenitors."			
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,			
RA	Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,			
RA	Nakamura Y., Nagahara K., Masuko Y., Sasaki N.;			
RT	"NEDO human cDNA sequencing project.";			
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.			
CC	-I- SUBUNIT: Interacts with HDAC3.			
CC	-I- SIMILARITY: Contains 1 Lish domain.			
CC	-I- SIMILARITY: Contains 8 WD repeats.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF314544; AAK0301.1; -			
DR	EMBL; AF268193; AAG44736.1; -			
DR	EMBL; AK022956; BAB14331.1; -			
DR	InterPro; IPR006594; LISH.			
DR	InterPro; IPR001680; WD40.			
DR	Pfam; PF00400; WD40; 8.			
DR	PRINTS; PR00320; GPROTEINBRPT.			
DR	ProDom; PD000018; WD40; 3.			
DR	SMART; SM00667; LISH; 1.			

Q09990 caenorhabdi
Q14775 homo sapien
P54314 mus musculu
Q93847 caenorhabdi
P49178 zea mays (m
Q09715 schizosacch
Q9uuc8 schizosacch
P04901 homo sapien
Q9fuy2 arabidopsis
P54311 rattus norv
Q40687 oryza sativ
Q22212 arabidopsis

34 299 10.9 665 1 LI23_CABEL
35 294.5 10.8 395 1 GBB5_HUMAN
36 292.5 10.7 395 1 GBB5_MOUSE
37 292 10.7 395 1 YZLL_CABEL
38 291 10.7 380 1 GBB_MAIZE
39 288.5 10.6 614 1 TU11_SCHPO
40 288 10.5 586 1 TU12_SCHPO
41 285.5 10.5 340 1 GBB1_HUMAN
42 285 10.4 931 1 LUG_ARATH
43 284.5 10.4 340 1 GBB1_RAT
44 283 10.4 380 1 GBB_ORYSA
45 283 10.4 554 1 PRP4_ARATH

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

[1]
NCBI_TaxID=9606;
SEQUENCE FROM N.A.
MEDLINE=9264241; PubMed=10330347;
Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,
Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.,
X-linked late-onset sensorineural deafness caused by a deletion
involving OAI and a novel gene containing WD-40 repeats.",
Am. J. Hum. Genet. 64:1604-1616 (1999).

[2]
SEQUENCE FROM N.A.
TISSUE=Lymph;
MEDLINE=22389257; PubMed=12477932;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin T.B., Toshivuki S., Carninci P., Mullany S.J.,
Rana S.S., Iqbalan N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Boak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Hahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.N., Krzywinski M.I., Skalska U., Smaluk D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- TISSUE SPECIFICITY: Ubiquitous
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).

EMBL; Y12781; CRA73319.1; ALT INIT.
EMBL; BC032708; AAH32708.1; --
DR MIM; 300196; --
DR GO; GO:0007605; P:hearing; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR GO; GO:0007601; P:vision; TAS.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PRO0320; GPOTHEINRPT.
DR Prodom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS50896; Lish; 1.
DR PROSITE; PS00578; WD_REPEATS_1; 4.
DR PROSITE; PS50882; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT DOMAIN 4 36 LISH.
FT REPEAT 179 218 WD 1.
FT REPEAT 235 274 WD 2.
FT REPEAT 276 315 WD 3.
FT REPEAT 318 358 WD 4.
FT REPEAT 359 398 WD 5.
FT REPEAT 401 449 WD 6.
FT REPEAT 452 491 WD 7.

SMART: SM00320; WD40; 8.
PROSITE; PS50896; LISH; 1.
PROSITE; PS00678; WD_REPEATS_1; 4.
PROSITE; PS50882; WD_REPEATS_2; 6.
PROSITE; PS50294; WD_REPEATS_REGION; 1.
Repeat; WD repeat.
DOMAIN 4 36 LISH.
REPEAT 167 206 WD 1.
REPEAT 223 262 WD 2.
REPEAT 264 303 WD 3.
REPEAT 306 344 WD 4.
REPEAT 347 386 WD 5.
REPEAT 389 427 WD 6.
REPEAT 440 479 WD 7.
REPEAT 481 513 WD 8.
DOMAIN 108 118 POLY-ALA.
E -> K (IN REF. 2).
Y -> H (IN REF. 3).
A -> Q (IN REF. 2).
CONFLICT 31 31
CONFLICT 59 59
CONFLICT 389 389
SEQUENCE 514 AA; 55595 MW; OB556D2EB4B796D CRC64;

Query Match 98.7%; Score 3696; DB 1; Length 514;
Best Local Similarity 98.8%; Pred. No. 5.7e-188;
Matches 508; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSISSEVNFLVRYLOSGSHSAFTGIESHSQSNGINGALVPPAALISIITKGLOYV 60
DB 1 MSISSEVNFLVRYLOSGSHSAFTGIESHSQSNGINGALVPPAALISIITKGLOYV 60
QY 61 EAIVSNEDGTLDGRPEIETSLIDAVMDVVQTRQAAYRDKLAAQAAAAAAAATNQ 120
DB 61 EAIVSNEDGTLDGRPEIETSLIDAVMDVVQTRQAAYRDKLAAQAAAAAAAASQ 120
QY 121 QGSAXNGENTANGENGATIANNTDMVEVDGVETPSNKAVALRGHESVFICAWNPV 180
DB 121 QGSAXNGENTANGENGATIANNTDMVEVDGVETIPPNKAVALRGHESVFICAWNPV 180
QY 181 SLLLVSGSDSTARWNLSNSTSGTQLVLRHCIREGGDVPNKDVTSLDWNSGTL 240
DB 181 SLLLVSGSDSTARWNLSNSTSGTQLVLRHCIREGGDVPNKDVTSLDWNSGTL 240
QY 241 ATGYSDGPARIWTCKGNLASTLGQHKGFIFALKWKNNKFILSAGVDKTTIIDAHTEA 300
DB 241 ATGYSDGPARIWTCKGNLASTLGQHKGFIFALKWKNNKFILSAGVDKTTIIDAHTEA 300
QY 301 KOQFPFHPAPALDVMQSNNTFASCSTDMCTHVCKLGQDRPKTFQGHTEVNAIKWDPT 360
DB 301 KOQFPFHPAPALDVMQSNNTFASCSTDMCTHVCKLGQDRPKTFQGHTEVNAIKWDPT 360
QY 361 GNLLASCSDDMTLKTKWSMKQDCNVHDLOAHKEIYTIKWSPTGPTGNPNNANMLASAF 420
DB 361 GNLLASCSDDMTLKTKWSMKQDCNVHDLOAHKEIYTIKWSPTGPTGNPNNANMLASAF 420
QY 421 DSTVRLWDVRGICHTLTLPKHQPVSVAFPSDGRYLASGSPDKCVHWNTQTGALVHSY 480
DB 421 DSTVRLWDVRGICHTLTLPKHQPVSVAFPSDGRYLASGSPDKCVHWNTQTGALVHSY 480
QY 481 RTGGGIFEVCWNAAGDKVGASADSGSCVCLDLRK 514
DB 481 RTGGGIFEVCWNAAGDKVGASADSGSCVCLDLRK 514

RESULT 2
TELX_HUMAN STANDARD; PRT; 526 AA.
AC AC 060907;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transducin beta-like 1X protein (Transducin-beta-like 1, x-linked).
GN TBLLX OR TBLL1.
OS Homo sapiens (Human).
OS Chordata; Vertebrata; Euteleostomi;

```
FT REPEAT 493 525 WD 8.
FT DOMAIN 108 124 POLY-ALA.
SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F6E9 CRC64;

Query Match
Best Local Similarity 89.3%; Score 2439; DB 1; Length 526;
Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

QY 1 MSISDEVNFLVRYLOESGFSHSAFTFGIESHISQSNGINGALVPPAALISIIQKGLQYV 60
Db 1 MSISDEVNFLVRYLOESGFSHSAFTFGIESHISQSNGINGALVPPAALISIIQKGLQYV 60
QY 61 EAEVINEGDTLFCGRPIESLSIIDAVMPDVVQTRQOAYRDKLAQOHAHAHAHAAT-- 118
Db 61 EAEVINEGDTLFCGRPIESLSIIDAVMPDVVQTRQOAYRDKLAQOHAHAHAHAAT-- 118
QY 119 -----NQOQSAKNGENTANGEENGAAHTIANHNTDMVEVDGDVEIPSNKAVLRG 167
Db 121 ATAATTSGVSHQNPKNREATVNGEENRAHSV-NNAKPEVIDGEVEIPSSKATVLRG 179
QY 168 HESEVFIQAWNPVSDLLVSGSDSTARIWNLSNSTSGPTOLVLRHCIREGGQDVPSNKG 227
Db 180 HESEVFIQAWNPVSDLLVSGSDSTARIWNLSNSTSGPTOLVLRHCIREGGQDVPSNKG 239
QY 228 VTSIDWNSSEGTLATGSDGFARITWDCGNLSTLQGHKGFIEALKNKNGNFILSAGVD 287
Db 240 VTSIDWNTGTLATGSDGFARITWDCGNLSTLQGHKGFIEALKNKNGNFILSAGVD 299
QY 288 KTTIWDHAHTGEAKQFPFHSAAPALVDVQSNNTFASCTDMCHIVCKLQGDRIKTFQG 347
Db 300 KTTIWDHAHTGEAKQFPFHSAAPALVDVQSNNTFASCTDMCHIVCKLQGDRIKTFQG 359
QY 348 HTNEVNAIKWPTGNLASCSDMTLKIWSMKQDCVHDLOAHNKEIYTIKWSPTGFTN 407
Db 360 HTNEVNAIKWPTGNLASCSDMTLKIWSMKQDCVHDLOAHNKEIYTIKWSPTGFTN 419
QY 408 NPANMLASASFDSTVRLWVDVDRGICHTLTKHEQEPVYSVAFSPDGRYLAGSFDKCVH 467
Db 420 NPANMLASASFDSTVRLWVDVDRGICHTLTKHEQEPVYSVAFSPDGRYLAGSFDKCVH 479
QY 468 IWNQTGALVHSYRGTFGIFVCNNAAGDKVGASASDGSVCVLDLRK 514
Db 480 IWNQTGALVHSYRGTFGIFVCNNAAGDKVGASASDGSVCVLDLRK 526

RESULT 3
ID TELX_HUMAN
AC Q9BQ87; STANDARD; PRT; 522 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transducin beta-like 1Y protein (Transducin-beta-like 1, Y-linked).
GN TBL1Y OR TBL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,
RA Wilson R.K., Waterston R.H., Page D.C.;
RT "The DNA sequence of the human Y chromosome."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC -----

or send an email to license@isb-sib.ch.
-----
CC ENBL; AF332220; AAK13472.1; -
CC ENBL; AF332221; AAK13473.1; -
CC ENBL; AF332222; AAK13474.1; -
CC Genew; HGNC:18502; TBL1Y.
CC InterPro; IPR006594; Lish.
CC Pfam; PF00400; WD40; 8.
CC PRINTS; PR00320; GPROTEINBRPT.
CC ProDom; PD000018; WD40; 3.
CC SMART; SM00667; Lish; 1.
CC SMART; SM00320; WD40; 8.
CC PROSITE; PS00896; Lish; 1.
CC PROSITE; PS00678; WD_REPEATS_1; 4.
CC PROSITE; PS00082; WD_REPEATS_2; 6.
CC PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT DOMAIN 4 36 LISH.
FT REPEAT 177 216 WD 1.
FT REPEAT 233 272 WD 2.
FT REPEAT 274 313 WD 3.
FT REPEAT 316 354 WD 4.
FT REPEAT 357 396 WD 5.
FT REPEAT 399 447 WD 6.
FT REPEAT 450 489 WD 7.
FT REPEAT 491 521 WD 8.
SQ SEQUENCE 522 AA; 56688 MW; 4E020216422442D8 CRC64;

Query Match
Best Local Similarity 86.1%; Score 2351; DB 1; Length 522;
Matches 439; Conservative 35; Mismatches 37; Indels 12; Gaps 2;

QY 1 MSISDEVNFLVRYLOESGFSHSAFTFGIESHISQSNGINGALVPPAALISIIQKGLQYV 60
Db 1 MSISDEVNFLVRYLOESGFSHSAFTFGIESHISQSNGINGALVPPAALISIIQKGLQYV 60
QY 61 EAEVINEGDTLFCGRPIESLSIIDAVMPDVVQTRQOAYRDKLAQOHAHAHAHAAT-- 115
Db 61 EAEVINEGDTLFCGRPIESLSIIDAVMPDVVQTRQOAYRDKLAQOHAHAHAHAAT-- 115
QY 116 -----AATNQOQSAKNGENTANGEENGAAHTIANHNTDMVEVDGDVEIPSNKAVLRGHE 169
Db 121 AATNQOQSAKNGENTANGEENGAAHTIANHNTDMVEVDGDVEIPSNKAVLRGHE 179
QY 170 SEVFIQAWNPVSDLLVSGSDSTARIWNLSNSTSGPTOLVLRHCIREGGQDVPSNKGVT 229
Db 180 SEVFIQAWNPVSDLLVSGSDSTARIWNLSNSTSGPTOLVLRHCIREGGQDVPSNKGVT 239
QY 230 SLDWNSSEGTLATGSDGFARITWDCGNLSTLQGHKGFIEALKNKNGNFILSAGVDKT 289
Db 240 SLDWNSDGTLTLAGSYDGFARITWDCGNLSTLQGHKGFIEALKNKNGNFILSAGVDKT 299
QY 290 TTIWDHAHTGEAKQFPFHSAAPALVDVQSNNTFASCTDMCHIVCKLQGDRIKTFQGH 349
Db 300 TTIWDHAHTGEAKQFPFHSAAPALVDVQSNNTFASCTDMCHIVCKLQGDRIKTFQGH 359
QY 350 NEVNAIKWPTGNLASCSDMTLKIWSMKQDCVHDLOAHNKEIYTIKWSPTGFTN 409
Db 360 NEVNAIKWPTGNLASCSDMTLKIWSMKQDCVHDLOAHNKEIYTIKWSPTGFTN 419
QY 410 NANMLASASFDSTVRLWVDVDRGICHTLTKHEQEPVYSVAFSPDGRYLAGSFDKCVHIW 469
Db 420 NSSIMLASASFDSTVRLWVDVDRGICHTLTKHEQEPVYSVAFSPDGRYLAGSFDKCVHIW 479
QY 470 NTCGTGALVHSYRGTFGIFVCNNAAGDKVGASASDGSVCVLDL 512
Db 480 NTCGTGALVHSYRGTFGIFVCNNAAGDKVGASASDGSVCVLDL 522

RESULT 4
ID TELX_MOUSE
AC Q9BQ87; STANDARD; PRT; 313 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transducin beta-like 1Y protein (Transducin-beta-like 1, Y-linked).
GN TBL1Y OR TBL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,
RA Wilson R.K., Waterston R.H., Page D.C.;
RT "The DNA sequence of the human Y chromosome."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC -----
```



```
DR PROSITE; PS50082; WD_REPEATS_2; 4.  
KW Repeat; WD_REPEATS_REGION; 1.  
FT DOMAIN 4 36  
FT REPEAT 163 208  
FT REPEAT 218 257  
FT REPEAT 259 298  
FT REPEAT 315 354  
FT REPEAT 337 396  
FT REPEAT 339 437  
FT REPEAT 440 479  
FT CONFLICT 396 396  
SQ SEQUENCE 535 AA; 59145 MW; 25DD19AB29FBAB07 CRC64;  
  
Query Match 18.4%; Score 501.5; DB 1; Length 535;  
Best Local Similarity 25.4%; Pred. No. 6.8e-29;  
Matches 143; Conservative 101; Mismatches 23; Indels 79; Gaps 16;  
  
QY 1 MSISDEYVNFVYRYLOESGFSHGAFTGIESHSQSIN-GALVPPAALSIIOKQLOY 59  
DB 1 MSITSEELNYLWRYCQMGHEVSALALQDTRVLEFDEKYEHIPLGLTVNLVQRGILY 60  
  
QY 60 VEAEVSIINEDG---TLFDGRPIESLSLIDAVMPVQTRQAYRDKLAQQAHAHAHAHA 116  
DB 61 TESELMVDSKGDISALNEHLSDFNLVQALQID-----KEKPE--ISSEGRFTL 109  
  
QY 117 ATNCGSAKNGENTANGENGAAHTIANNHTDM-EVGDVPIPSNKAVVLRGHSEVFIC 175  
DB 110 ETNSE-SNKAEDGASVETRTQEDDTSIDSSDLDGFKI--LKSIV---KLDNIYSS 163  
  
QY 176 AMNPVSD-LLVSGSGDSTARIWLSNSENSTGFTQ-----LVLRHCIREGQDVPSNKDV 228  
DB 164 TWPPLDESILAYGEKNSVARLARIVETDQEKYKWLTIIAELRHPPALSASSGKTIQV 223  
  
QY 229 TSLDWNSEGLTATSGVDGFARIWTKDGNLSTLQGHKPIFALKWKNKGNFILSAGVDK 288  
DB 224 TCLAWSDGNSIVTGVENGELRLNKTGALLNLNFRHAPIVSVKWKNDGTHIISMDVEN 283  
  
QY 289 TTIWDAHTGEAKQOFPF-----HSAP-----ALDWDQSNNTFASCSGSDMCIHV 333  
DB 284 VTILNNVISGVMGHFELKETGGSSINAEHSGSLGVDVVEWDDKFIPIPGKGAIFV 343  
  
QY 334 CKLGQDRPIKTFQGHTEVNAIKWDPGNLNASGSDMTLKIMSKMDCNVDHDLQAHNKE 393  
DB 344 YQITEKTPGKLIGHGPIISVLEFNDINKLLSASDDGTLRWHGNGNSQNSFYGHQS 403  
  
QY 394 IYTIKWSPTGPTNPNANMLASGDSVRLVDVDRGICIHLLTKHQEPVYSVAESP 453  
DB 404 IVSASWV-----GDDKVISCSMDGVSRLMSLKNLTLALSIVDGVPIFAGRISQD 453  
  
QY 454 GRYLASGDFDKCVHIW-----NTQTGAL-----VHSYRGTGIGIFEVC 490  
DB 454 GQKVAFAFMQGVNVYDLKLNLSKRSRLYGNRDGILNPLPIPLVYASQSSQDNDYIFDLS 513  
  
QY 491 WNAAGDKVGA--SASDGSVCVL 510  
DB 514 WNCAGNKISVAYSLOEQGSVVAI 535  
  
RESULT 6  
YY46 ANASP  
ID YY46 ANASP STANDARD; PRT; 1526 AA.  
AC Q8YRII;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical WD-repeat protein alr3466.  
GN ALR3466  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.
```

MEDLINE=21595285; PubMed=11759840;
Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
Watanabe A., Itiguchi M., Ishikawa A., Kawashima K., Kimura T.,
Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
DNA Res. 8:205-213(2001).
CC -|- SIMILARITY: Contains 16 WD repeats.
CC -|- SIMILARITY: Contains 1 pentapeptide repeat domain.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; AP003593; BAB75165.1; -;
DR PIR; AC2239; AC2239.
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 15.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD000018; WD40; 8.
DR SMART; SMC0320; WD40; 15.
DR PROSITE; PS00678; WD_REPEATS_1; 11.
DR PROSITE; PS50082; WD_REPEATS_2; 15.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
DR PROSITE; PS50837; NACHT; UNKNOWN 1.
KW Hypothetical protein; Repeat; WD_repeat; Complete proteome.
FT REPEAT 334 376
FT DOMAIN 823 862
FT REPEAT 862 901
FT REPEAT 904 945
FT REPEAT 946 985
FT REPEAT 988 1027
FT REPEAT 1030 1069
FT REPEAT 1072 1111
FT REPEAT 1114 1153
FT REPEAT 1156 1195
FT REPEAT 1198 1237
FT REPEAT 1240 1279
FT REPEAT 1282 1321
FT REPEAT 1324 1363
FT REPEAT 1366 1405
FT REPEAT 1408 1447
FT REPEAT 1450 1491
SQ SEQUENCE 1526 AA; 170376 MW; D5DF272509B4A738 CRC64;

Query Match 17.6%; Score 481; DB 1; Length 1526;
Best Local Similarity 30.3%; Pred. No. 8.6e-27;
Matches 118; Conservative 77; Mismatches 149; Indels 46; Gaps 11;

QY 136 NGAHTIANNHTDMMEVDGVEIPSNKAV-VLRGHESEVFICAWNPNVSDLLVSGSGDSTAR 194
DB 1127 NGV-TLANGSSDQIVRLWD--ISSKKCLYTLQHTNWNNAVAFSPDGTALSSGSDQIVR 1183

QY 195 IWNLSNENSTGPTQLVLRHC--IRGGQGVDPNKNVTSLDWNSGTLATGSDGFARIW 252
DB 1184 LWDISSK-----CLYLQG-----HTSVNVSVFNPDGSLASGSDQIVRLW 1227

QY 253 -TKDGNLSTLGHKPIFALKWKNKGNFILSAGVDKTTIWDHTGEAKQOFPFHSAPA 311
DB 1228 EINSKCLCTFPQHTSWNSVFNPDGSLASGSDKTVRLWDISSKCLHTFQHT--- 1284

QY 312 LDVDWQSNNTF-----ASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPGNLL 364
DB 1285 ---NWVNSVAFNPDGSLASGSDQIVRLWEISSKCLHTFQHTSWVSVTFSPDGTML 1341

QY 365 ASCSDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNPNANMLASGSDIV 424

1342 ASGSDQTVRLWSISSGECYTLFLGHTNMGWSVIFSPDG-----AIIASGSGDQTV 1392
425 RLWDVRCICHTLTKHQEPYVAFSPDGRYLASGDFDKVCHVNTQTGALVHVRG-T 483
1393 RLWISSGKCLYTLQGHNNMGWSVIFSPDGTLLASGSDQTVRLWNISSECLYTLHGHI 1452
484 GGIFPCVWNAAGDKVGSASDGSCVCLDLR 513
1453 NSVRVAFSSDGLIILASGSDDETIKLMDVK 1482
RESULT 7
HET1 PODAN
ID HET1 PODAN STANDARD; PRT; 1356 AA.
AC Q00808;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vegetative incompatibility protein HET-E-1.
GN HET-E1.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96009891; PubMed=7557402;
RA Saube S., Turcq B., Begueret J.;
RT "A gene responsible for vegetative incompatibility in the fungus
Podospora anserina encodes a protein with a GTP-binding motif and G
beta homologous domain."
RL Gene 162:135-139(1995).
CC -!- FUNCTION: RESPONSIBLE FOR VEGETATIVE INCOMPATIBILITY THROUGH
CC SPECIFIC INTERACTIONS WITH DIFFERENT ALLELES OF THE UNLINKED GENE,
CC HET-C.
CC -!- SIMILARITY: Contains 10 WD repeats.
CC -!- SIMILARITY: Contains 1 NACHT domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L28125; AAA85775.1; -
DR PIR; T18521;
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF05729; NACHT; 1.
DR Pfam; PF00400; WD40; 10.
DR PRINTS; P000320; GPROTEINBRPT.
DR ProDom; P0000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PS00837; NACHT; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 10.
DR PROSITE; PS00082; WD_REPEATS_2; 10.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW GTP-binding; Repeat; WD repeat.
FT DOMAIN 294 629
FT NP BIND 307 869
FT REPEAT 839 869
FT REPEAT 881 911
FT REPEAT 923 953
FT REPEAT 965 995
FT REPEAT 1007 1037
FT REPEAT 1049 1079
FT REPEAT 1091 1121
FT REPEAT 1133 1163
FT REPEAT 1175 1205
FT REPEAT 1217 1247

SQ SEQUENCE 1356 AA; 149765 MW; 965FB319844E0651 CRC64;
Query Match 16.9%; Score 462; DB 1; Length 1356;
Best Local Similarity 27.3%; Pred. NO. 1.7e-25;
Matches 121; Conservative 74; Mismatches 181; Indels 68; Gaps 11;
QY 78 IESLSLIDAVMPDVVTRCQAVRDKLAQCHAAHAAAAAATNQGSAGNGB-----NT 130
DB 820 ISTISVVEAEMWACTQT-----LEGHGSSVLSVAFSADGQGVASGSDDKIKIWD 870
QY 131 ANGEENGAHTIANHNTDMMEVDGVEIIPSNKAVLRLGHESEVFCAMNPVSDLLVSGSGD 190
DB 871 ASG--TGTQT-----LEGHGGSVWSVAFSPDRERVASGSD 904
QY 191 STARIWNLSENSTGPTQLVLRHCIREGGQDVFSNKQVTSIDMNSSEGLLATGSDGFAR 250
DB 905 KTIKIMDAASGTC---TOTLEGH-----VQSVAFSPDGQGVASGSDDHNIK 950
QY 251 IW-TKGNLNLASTLGOHKGPFAFKWKNKGNFILSAGVDKTTIINDAHTGEAKQQPFHSA 309
DB 951 IWDAASGTCQTLEHGSSVLSVAFSPDQQRVASGSGDKTIKIMDTASGTCQTLEHGSG 1010
QY 310 PALDVDMQ--SNNTFASCSTDMCIHVCKLQDRIPIKTFQHTNEVNAIKWDPGNLLASCS 368
DB 1011 SVMSVAFSPDQQRVASGSGDKTIKIMDTASGTCQTLEHGSGVSVVFSFSPDQQRVASGS 1070
QY 369 DDMTLKISMKQDNCVHDLQAHNKEIYTIKWSPTGPTNPNANLMLASASFDSTVRLWD 428
DB 1071 DHTIKINDAVSGTCQTLEHGSDSVWSVAFSPDQ-----RVASGSDIGIKIWD 1121
QY 429 VDRGICHTLTKHQEPYVAFSPDGRYLAGSFKCVKHIWNTQTGALVHVRGCG-IF 487
DB 1122 AASGTCQTLEHGSGVWSVAFSPDQQRVASGSDGTTIKINDAASGTCQTLEHGSGVWQ 1181
QY 488 EVCVNAAGKVGASASDGSCVCLVD 511
DB 1182 SVAFSPDQQRVASGSDDKTIKIMD 1205
RESULT 8
YS00 ANASP
ID YS00 ANASP STANDARD; PRT; 1258 AA.
AC Q8YTC2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical WD-repeat protein alr2800.
GN Alr2800.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- SIMILARITY: Contains 15 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

```

DR EMBL; AP003590; BAB74499.1; -
DR PIR; A12155; A12155.
DR InterPro; IPR002182; NB-ARC.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00931; NB-ARC; 1.
DR Pfam; PF00400; WD40; 14.
DR PRINTS; PR00320; GPROTEINERPT.
DR PRODOM; PD000018; WD40; 1.
DR SMART; SM00320; WD40; 14.
DR PROSITE; PS00678; WD REPEATS 1; 9.
DR PROSITE; PS0082; WD REPEATS 2; 14.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 55 93
FT REPEAT 640 679
FT REPEAT 682 721
FT REPEAT 724 763
FT REPEAT 766 807
FT REPEAT 809 849
FT REPEAT 850 889
FT REPEAT 892 931
FT REPEAT 934 975
FT REPEAT 976 1017
FT REPEAT 1019 1059
FT REPEAT 1060 1101
FT REPEAT 1103 1143
FT REPEAT 1144 1183
FT REPEAT 1186 1227
SQ SEQUENCE 1258 AA; 139513 NW; 45DF03B91170C451 CRC64;

Query Match 16.8%; Score 459.5; DB 1; Length 1258;
Best Local Similarity 31.5%; Pred. No. 2.4e-20; Indels 77; Gaps 7;
Matches 102; Conservative 44; Mismatches 101;

QY 165 LRGHSEVFICAMNPVSDLLVSGSDSTARIWNLSENSTSGTQLVLRHCIREGGQDVPS 224
DB 974 LEGHTDFIYGLAFSPDSQTLASASTDSSVLWNI---STGCFCILLEH----- 1019
QY 225 NKDVTSLDW-----NSEGITLATGSDGFGARWT-KDGNLASTLGQHKGPALKWKK 277
DB 1020 -----TDWYAVVVFHFGKIIATGSDCTVKLWNISTGQCLTSLSHSKILGMASPD 1073
QY 278 GNFLSAGVDKTTIWDHAHTGEAKQPFHSPALPDVDVDSNNTFASCSTDMCHIVCKLG 337
DB 1074 GOLLASASQSVRLWD-----CCTGRCVGI----- 1099
QY 338 QDRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWMKQDNCVHDLQAHNKEIYTI 397
DB 1100 -----LRGHSNRVYSAIFSPNGEIIATGSDTQTVXINDWQOQCKCLKTLTGHTNWVFI 1152
QY 398 KWSPTGPTNPNANMLASASDSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGYL 457
DB 1153 AFSPDGK-----ILASASHDQTVRIWDVNFVKCHHICIGHTHLVSSVAFSPDGEV 1203
QY 458 AGSFDDKCVHIWNTQTGALVHSYR 481
DB 1204 AGSGDQTVRIWNVKTGECILQLR 1227

RESULT 9
YL24 ANASP STANDARD; PRT; 1683 AA.
AC Q8YV57;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein all2124.
GN ALL2124.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxId=103690;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RL Cyano bacterium Anabaena sp. strain PCC 7120.";
CC -!- SIMILARITY: Contains 13 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003598; BAB73823.1; -
CC PIR; AF2071; AF2071.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 14.
CC PRINTS; PR00320; GPROTEINERPT.
CC SMART; SM00320; WD40; 14.
CC PROSITE; PS00678; WD REPEATS 1; 3.
CC PROSITE; PS0082; WD REPEATS 2; 14.
CC PROSITE; PS0294; WD REPEATS REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1070 1109
FT REPEAT 1111 1150
FT REPEAT 1153 1192
FT REPEAT 1195 1236
FT REPEAT 1280 1318
FT REPEAT 1320 1359
FT REPEAT 1361 1400
FT REPEAT 1402 1441
FT REPEAT 1444 1483
FT REPEAT 1486 1523
FT REPEAT 1528 1567
FT REPEAT 1570 1609
FT REPEAT 1611 1650
SQ SEQUENCE 1683 AA; 185459 MW; F7CB361FF54F7137 CRC64;

Query Match 16.5%; Score 451; DB 1; Length 1683;
Best Local Similarity 28.5%; Pred. No. 1.5e-24;
Matches 109; Conservative 71; Mismatches 163; Indels 40; Gaps 7;

QY 124 AKNGENTANGENGAAHTIANNHTDMMEYDGDVEIPSKAVVLRGHESEVFICAMNPVSD 183
DB 1290 SSDGRAIASASRDNTIKLWNRHGLE-----TFTGSGGVYAVNPLDSNI 1336
QY 184 LVSGSGDSTARIWNLSENSTSGTQLVLRHCIREGGQDVPSNKVDTSLDNSEGTLATG 243
DB 1337 IASASLDNTIRLW---QRPLISPLEV-----LAGNSGVYAVSFLHDGSIATA 1381
QY 244 SYDGFARIW-TKDGNLASTLGQHKGPALKWKNKGNFLSAGVDKTTIWDHAHTGEAKQ 302
DB 1382 GADGNIQLWHSQDGSLLKTLPGNKA-IYGISFTPGDILASANADTKVINEVRCKALK 1440
QY 303 QPFFHSAPALVDWQ-SNNTFASCSTDMCHIVCKLGQDRPIKTFQGHTEVNAIKWDPTG 361
DB 1441 TLIGHDNEVNVKVPFSPDGTQLASASRDNTVKLWNVSDGKFKTKLGHTEDEVFWVSFSDG 1500
QY 362 NLLASCSDDMTLKIWMKQDNCVHDLQAHNKEIYTIKWSPTGPTGNPNANMLASASPD 421
DB 1501 KIIASASADTKIRLWDSFGNLIKSLPAHNDLVYSVNFNPDGS-----MLASTSAD 1551
QY 422 STVRLWDVDRGICHTLTKHQEPVYVAFSPDGYLASGSPDKCVHIWNTQTGALVHSYR 481
DB 1552 KTVKLWRSHDGHLLHTFSGHSNVVYSSFSFSPDGRVIASASEDKTVKIWDIGHLLTLPQ 1611
QY 482 GTGGIFEVCWNAAGDKVGASASD 504

```



```

RL Mol. Biol. Cell 13:4100-4109(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK001320; BAA91621.1; -
DR EMBL; BC002884; AAH02884.2; -
DR EMBL; BC02075; AAH02075.1; -
DR EMBL; AJ005257; CAA06444.1; -
DR SWISS-2DPAGE; Q9NVX2; HUMAN.
DR InterPro; IPR001632; Gpotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS_1; 4.
DR PROSITE; PS00678; WD REPEATS_2; 7.
DR PROSITE; PS50082; WD REPEATS_REGION; 1.
DR PROSITE; PS50294; WD REPEATS_REGION; 1.
KW Nuclear protein; Repeat; WD Repeat.
FT REPEAT 112 151
FT REPEAT 151 193
FT REPEAT 193 241
FT REPEAT 241 282
FT REPEAT 282 326
FT REPEAT 326 366
FT REPEAT 366 409
FT REPEAT 409 451
FT REPEAT 451 484
FT REPEAT 484 532
SQ SEQUENCE 485 AA; 53266 MW; 0AEBE24B44957379 CRC64;
Query Match 13.3%; Score 363.5; DB 1; Length 485;
Best Local Similarity 26.3%; Pred. No. 6e-19;
Matches 102; Conservative 47; Mismatches 146; Indels 89; Gaps 9;
QY 165 LRGHSEVFICAWNPVSDLLVSGSGDSTARIWNLSEN-----STSGPTQLVLRHCIRGGQ 220
DB 110 LEGHSEAVISAFSPGTYGKYLASGSGDTTVPFWDLSLTETPFTCKGRHFWLISGWSPDGK 169
QY 221 DVPS-----NKGDTSLDW-----NSEGTLATGSDYDGFAR 250
DB 170 KLASCKKGQILLDPSTGKQVGRNLGSKWITGLSWEPLHNPCEYVASSSKDGSVR 229
QY 251 IW-TKDGNLASTLGQHGPIFALKWKNKGNFILSGVDKTTIWDATGGEAKQFPFH-- 307
DB 230 IWDTTAGRCERILTGTQSVTCRLRGWGG-LLYSASQDRTIKVRAHDGVLCTLQGHG 288
QY 308 --SAPALVDW-----QSNNTFASCS 326
DB 289 WYNTWALSTDYALRTCAFEPAEASVNPDOGLSQELKERALSYNLVRGGGPERLVSGS 348
QY 327 TDMCIHVCKLQGD-RPIKTFQGHTEVNAIKWDPTGNLLASCSDMTLKIWSKQDNCVH 385
DB 349 DDFTLFLSPAEDKKPLRTMTGHOALINQVLPSPDSRIVASASFDKSIKLDGRTGKYLA 408
QY 386 DLQAHNKIYTIKWSTGTGNTNPNANLMLASASFDSTVRLWDVDRGICHTLTKHQPVP 445
DB 409 SLRGHVAAYQIAWSAD-----SRLLVSGSGSDSTLKVDWVKAQKLAWDLPGHADEV 459
QY 446 YSAFSPDGRYLASGDFKCVHIW 469
DB 460 YAVDMSPDQRVASGGKDKCLRIW 483

```

RESULT 13
HUS7_MOUSE

```

ID HUS7_MOUSE STANDARD; PRT; 485 AA.
AC Q8VEI4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE WD-repeat protein HUSY-07.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lomellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; BC018399; AAH18399.1; -
DR InterPro; IPR001632; Gpotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS_1; 4.
DR PROSITE; PS00678; WD REPEATS_2; 7.
DR PROSITE; PS50082; WD REPEATS_REGION; 1.
DR PROSITE; PS50294; WD REPEATS_REGION; 1.
KW Nuclear protein; Repeat; WD Repeat.
FT REPEAT 112 151
FT REPEAT 151 193
FT REPEAT 193 241
FT REPEAT 241 282
FT REPEAT 282 326
FT REPEAT 326 366
FT REPEAT 366 409
FT REPEAT 409 451
FT REPEAT 451 484
FT REPEAT 484 532
SQ SEQUENCE 485 AA; 53148 MW; 2D7F59D603AEC07B CRC64;
Query Match 13.3%; Score 362.5; DB 1; Length 485;
Best Local Similarity 26.3%; Pred. No. 7.1e-19;
Matches 101; Conservative 48; Mismatches 146; Indels 89; Gaps 9;
QY 165 LRGHSEVFICAWNPVSDLLVSGSGDSTARIWNLSEN-----STSGPTQLVLRHCIRGGQ 220

```

Db 110 LEHSEAVISVAPSPGKYLASGSDTTRVFWDLSTETPHFTCKGHRHWLISWSPDGK 169
Qy 221 DVPS-----NKDVTSLDW-----NSETLLATSGYDGFAR 250
Db 170 KLASGCKNGQILLWDPSTGLQVGRTLGHSKWIITGLSWEPFLHNPCECRYVASSKDGSVR 229
Qy 251 IW-TKDGNLASTLGHKGPPIFALKWKNKGNFILSAGVDKTTIWDATHTGEAKQFPPH-- 307
Db 230 VWDITAGRCERILTGHTQSVTCLEWGGDG-LLISASQDRTIKVRWRAHDGVLCTLQGHG 288
Qy 308 --SAPALDVM-----QSNNTFASCS 326
Db 289 WNTWALSTDYALRTGAFAPAEATVNAQDLQGLSKELKERASSRYNLVRGQPERLVSGS 348
Qy 327 TDMCIHVKCLGQD-RPIKTFQGHTEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQDNCVH 385
Db 349 DDFTLFWSPADPKPIARMTGHQALNQVLFSPDSRIVASASPKSIKLWDGRTGKYLA 408
Qy 386 DLQAHNKEIYTKWSPGTGNTNPNANMLASAFSDTVLMDVDRGICHTLTKHOEPV 445
Db 409 SLRGHVAAYQIASAD-----SRLLVSGSDSTLKVWQVYKQKLAITDLPGHADV 459
Qy 446 YSVAFSPDGRYLAGSPDKCVHIW'459
Db 460 YAVDWSPDGRVAGSGDKCLRHW 483

RESULT 14

AAC3_DICDI
ID -AAC3_DICDI STANDARD; PRT; 437 AA.
AC P14197;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE AAC-rich mRNA clone AAC3 protein (Fragment).
GN AAC3.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90066348; PubMed=2511421;
RA Shaw D.R., Richter H., Giorda R., Ohmachi T., Ennis H.L.;
RT "Nucleotide sequences of Dictyostelium discoideum developmentally
RT regulated cDNAs rich in (AAC) imply proteins that contain clusters of
RT asparagine, glutamine, or threonine."
RL Mol. Gen. Genet. 218:453-459 (1989).
CC -!- DEVELOPMENTAL STAGE: The concentration of AAC-rich mRNAs is low
CC in dormant spores and growing cells, but increases during
CC spore-germination and multicellular development.
CC -!- MISCELLANEOUS: Several proteins derive from AAC-rich mRNA, which,
CC due to a frameshift also have ACA and CAA codons and thus are
CC Asn-, Thr- or Gln-rich.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).
CC
CC EMBL; X16524; CAA34531.1; -.
DR PIR; S05357; S05357.
DR DictyBase; DDB0001891; AAC3.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 5.
DR PRINTS; PR00320; GPROTINBRPT.
DR SMART; SMC0320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00682; WD_REPEATS_2; 3.
DR PROSITE; PS02094; WD_REPEATS_REGION; 1.

KW Repeat; WD repeat.
FT NON_TER 1
FT REPEAT 122 161 WD 1.
FT REPEAT 185 227 WD 2.
FT REPEAT 229 266 WD 3.
FT REPEAT 269 308 WD 4.
FT REPEAT 316 355 WD 5.
FT REPEAT 358 397 WD 6.
FT REPEAT 399 436 WD 7.
FT DOMAIN 8 75 GLN-RICH.
SQ SEQUENCE 437 AA; 49151 MW; 755AE61228D0A1A9 CRC64;

Query Match 13.1%; Score 358; DB 1; Length 437;
Best Local Similarity 23.7%; Pred. No. 1.3e-18;
Matches 109; Conservative 79; Mismatches 175; Indels 96; Gaps 13;

Qy 93 QTRQAYRDKLAQQHAAAAAATNQGSAKNGENTANGEEGAHTIANHHTDMEVD 152
Db 28 QTQVOQLHNLQHQQNQIQQQAQATQCHLOTQYQLSQIHOQSQSQSLNN----- 79
Qy 153 GDVEIPSNKAVVLRGHESEVFICAWNPVSDLLVSGSGDSTARIWNLSENST-----SGP 206
Db 80 -----LNNSKEST-----NIPKINTOYTFNDEKN 104
Qy 207 TOLVLRHCIREGGQD-VPSNKDVTSLDMNSECTLLATGSDYGFARIW----- 252
Db 105 LDLASRYFSECTKDFIGNKKKSTVAVNANGTKIASSGSDGIVRVWNPDPGLGNSNNNN 164
Qy 253 -----TKDGNLST--LGQHKGPFAFKWKNKGNFIL-SAGVDKTTIWDATHTGEAKQ 302
Db 165 SNTTSNSKNNNIKETIELKGDHGSIEKISPKNNNDLLASAGTDKVIKWIDVKIGKIG 224
Qy 303 QPFFHSAPALDWDQSNNTF-ASCSTDMCIHVCKLQGRPIKTFQGHTEVNAIKWDPGTG 361
Db 225 TVSTNS-ENIDVRWSPDQGFIVACTRDDHALIDPTIKTLAIYKFNGBELAQVGDWNG 283
Qy 362 NLLASCSDDMTLKWSMKQDNCVH-----DLQAHNKEIYTKWSPGTGPTNPNANMLA 416
Db 284 DLILMANSNGNIEAYKFLPKSTTHVKHLTKLTYGHTASYICMEFDPTGK-----YLA 334
Qy 417 SASFDSVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSGFDKCVKIWN'TQTGAL 476
Db 335 AGSASISVSLWIEDMMCMVKTFIKSTFFCRSVSFSFDGQFIASSFESTIEFIHIESOP 394
Qy 477 VHSYRGTGIFVWCNV-----AAGDKV GASASDGSVCV 509
Db 395 IHTIE-CISGVSSLMWHPTLPLLAYAPEINNNKDPISIRV 432

RESULT 15

CIAL_HUMAN
ID CIAL_HUMAN STANDARD; PRT; 339 AA.
AC O76071;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE WD-repeat containing protein Cial 1.
GN CIAO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=98225157; PubMed=9556563;
RA Johnstone R.W., Wang J., Tommerup N., Vissing H., Roberts T., Shi Y.;
RT "Ciao 1 is a novel WD40 protein that interacts with the tumor
RT suppressor protein wt1."
RL J. Biol. Chem. 273:10880-10887 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Adams M.D., Loftus B.J., Zhou L., Crosby M., Fuhrmann J., Mason T.M.,
RA Brandon R., Kim U.J., Kerlavage A.R., Venter J.C.;

Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 (3)
 RN SEQUENCE FROM N.A.
 RP TISSUE-Blood, and Skin;
 RC MEDLINE=22388257; PubMed=12477932;
 RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Statchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences."
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=99367400; PubMed=10438340;
 RA Johnstone R.W., Tommerup N., Hansen C., Vissing H., Shi Y.;
 RT "Structural organization, tissue expression, and chromosomal
 localization of Ciao 1, a functional modulator of the Wilms' tumor
 suppressor, Wt1."
 RT Immunogenetics 49:900-905(1999).
 CC -!- FUNCTION: Seems to specifically modulate the transactivation
 activity of Wt1.
 CC -!- SUBUNIT: Interacts with Wt1.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- MISCELLANEOUS: 'Ciao' means 'bridge' in Chinese.
 CC -!- SIMILARITY: Contains 7 WD repeats.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; U63810; AAC24948.1; ..
 EMBL; AC004020; AAC23493.1; ..
 EMBL; BC001395; AAH01395.1; ..
 EMBL; BC032812; AAH32812.1; ..
 MIM; 604333; ..
 DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
 DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . . ; TAS.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 7.
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000018; WD40; 2.
 DR SMART; SM00320; WD40; 7.
 DR PROSITE; PS00678; WD REPEATS_1; 1.
 DR PROSITE; PS00682; WD REPEATS_2; 6.
 DR PROSITE; PS0294; WD REPEATS_REGION; 1.
 KW Repeat; WD repeat; Nuclear protein.
 FT REPEAT 14 53 WD 1.
 FT REPEAT 59 98 WD 2.
 FT REPEAT 103 142 WD 3.
 FT REPEAT 148 187 WD 4.
 FT REPEAT 192 231 WD 5.
 FT REPEAT 250 289 WD 6.
 FT REPEAT 301 338 WD 7.
 SEQUENCE 339 AA; 37840 MW; 63A8D8257A204FC8 CRC64;

Query Match 12.8%; Score 349; DB 1; Length 339;
 Best Local Similarity 30.4%; Pred. No. 4.2e-18;
 Matches 92; Conservative 43; Mismatches 88; Indels 80; Gaps 11;
 QY 222 VPSNKDVTS--LDWNSEGTLLATGSYDGFARITWTKDGN---LASTLGO-HKGPIFALKWN 275
 DB 11 VPAHPSRCWFLANPAGTLLASCGRIRIRINGTEGDSWICKSVLSGHQRTVRKVAWS 70
 QY 276 KXGNFILSAGVDKTTIINDAHTGEAKQQPPFHSAPALDVQSNNTFASCTDMCIHVCK 335
 DB 71 PCGNYLASASPDATTCTI-----WKKNQDDFECVT----- 99
 QY 336 LQQDRPIKTFQOHTNEVNAIKWDPTGNLLASGDDMTLKIKSMKQDN---CVHDIQAHNK 392
 DB 100 -----TLEGHENEVKSVAWAPSGNLLATCSRDKSVWVWEDEDEYEYCVSLNHTQ 151
 QY 393 EIVTTIKWSPGTGTTNNPNANMLASASFDSTVRLW--DVDRGICIHITLTKHQEPVYSVAF 450
 DB 152 DVKHVVWHP-----SQELLASASYDDTVKLYREEDDDWVCCATLEGHSTVWSLAF 202
 QY 451 SPDGRYLASGSPDKCVHIWNTOTGALVHSYRGTTGIFEVCAAGDKYGAASDGS---V 507
 DB 203 DFGGRLASCSDDRTVRIWRYLPG-----NEQG--VACSGSDPSWKCI 244
 QY 508 CVL 510
 DB 245 CTL 247

Search completed: August 9, 2004, 16:46:13
 Job time : 9.66667 secs

Result No.	Score	Query		DB	ID	Description
		Match	length			
1	2691	98.5	514	4	AAb95225	Human pro
2	2688	98.4	514	6	AAb007190	Human p53
3	2439	89.3	577	7	ADb14051	Human src
4	2342	85.8	542	5	ABp41760	Human ova
5	2027	74.2	395	5	ABp51424	Human MDD
6	1906.5	69.8	700	4	ABb60376	Drosophil
7	1891.5	69.3	584	4	ABG21351	Novel hum
8	1142	40.2	208	4	AAc04385	Human pol
9	824	30.2	167	4	ABG21350	Novel hum
10	498.5	18.3	535	6	ABR52930	Protein s
11	375.5	13.7	515	6	ABt53774	Protein s
12	374.5	13.7	414	6	ADa13321	Human int
13	374.5	13.7	514	2	AAr85881	WD-40 dom
14	366.5	13.4	481	4	ABb59486	Drosophil
15	365	13.4	411	4	ABb62260	Drosophil
16	365	13.4	479	3	AAy79678	Drosophil
17	364.5	13.3	485	4	AAb68284	Amino aci
18	363.5	13.3	485	4	AAb68516	Human GTP
19	363.5	13.3	485	4	AAb92844	Human pro
20	363.5	13.3	485	5	ABb97306	Human hum
21	356.5	13.1	484	4	AAb68282	Amino aci
22	355	13.0	352	3	AAg14893	Arabidops
23	355	13.0	352	3	AAg48119	Arabidops
24	355	13.0	352	7	ABb95040	A. thalia
25	351	12.9	349	3	AAg48120	Arabidops

QY 181 SLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGQDVPNSKDVTSLDWNSGELL 240
 DB 181 SLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGQDVPNSKDVTSLDWNSGELL 240
 QY 241 ATGSYDGFARLWTKDGNLSTLGHKGPIFALKWKKGNFLLSAGVDKTTIIMDAHTGEA 300
 DB 241 ATGSYDGFARLWTKDGNLSTLGHKGPIFALKWKKGNFLLSAGVDKTTIIMDAHTGEA 300
 QY 301 KQOFPFHSAAPALVDWQSNNTFASCTDMCIHVCKLQGDREIKTFQGHTEVNAIKWDPT 360
 DB 301 KQOFPFHSAAPALVDWQSNNTFASCTDMCIHVCKLQGDREIKTFQGHTEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKEIYTIKWSPTGCTGNPNANILMASAF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKEIYTIKWSPTGCTGNPNANILMASAF 420
 QY 421 DSTVRLWDVDRGICHTLTKEHPEVYSVAFSPDGRYLASGSDKCVHIMNTQTGALVHSY 480
 DB 421 DSTVRLWDVDRGICHTLTKEHPEVYSVAFSPDGRYLASGSDKCVHIMNTQTGALVHSY 480
 QY 481 RGTGGIFVVCWNAAGDKVGSASDGSVCVLDLRK 514
 DB 481 RGTGGIFVVCWNAAGDKVGSASDGSVCVLDLRK 514

RESULT 3

ADD14051
 ID ADD14051 standard; protein; 577 AA.
 XX AC ADD14051;

XX 01-JAN-2004 (first entry)

DE Human src biomarker polypeptide SEQ ID NO:240.

XX predictor set; protein tyrosine kinase activity modulator;
 KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
 KW gene therapy; drug sensitivity; genetic profile; cancer; human.
 XX Homo sapiens.

XX WO20003062395-A2.

XX 31-JUL-2003.

XX 17-JAN-2003; 2003WO-US001991.

XX 18-JAN-2002; 2002US-0350061P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Huang F, Fairchild CR, Lee FY, Shaw P;

XX WPI; 2003-636735/60.

XX N-PSDB; ADD14646.

XX New polynucleotides and polypeptides for predicting the activity of
 PT compounds that interact with protein tyrosine kinases and/or protein
 PT tyrosine kinase pathways.

XX Claim 10; SEQ ID NO 240; 139pp; English.

XX The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and
 CC correlating the expression of the markers to the compound's ability to
 CC modulate the activity of the cells; (2) a plurality of cell lines for
 CC identifying polynucleotides and polypeptides whose expression levels

CC correlate with compound sensitivity or resistance of cells associated
 CC with a disease state; and (3) identifying polynucleotides and
 CC polypeptides that predict compound sensitivity or resistance of cells
 CC associated with a disease state, comprising subjecting the plurality of
 CC cell lines to one or more compounds, analysing the expression pattern of
 CC a microarray of polynucleotides or polypeptides, and selecting
 CC polynucleotides or polypeptides that predict the sensitivity or
 CC resistance of cells associated with a disease state by using the
 CC expression pattern of the microarray. The polynucleotides and
 CC polypeptides have cytostatic activities, and can be used in gene therapy.
 CC The polynucleotides and polypeptides are useful in predicting the
 CC activity of compounds that interact with protein tyrosine kinases and/or
 CC protein tyrosine kinase pathways. These may be used in determining drug
 CC sensitivity in patients to allow the development of individualized
 CC genetic profiles which aid in treating diseases and disorders (e.g.
 CC cancer) based on patient response at a molecular level. The present
 CC sequence is used in the exemplification of the present invention.

XX SQ Sequence 577 AA;

Query Match 89.3%; Score 2439; DB 7; Length 577;

Best Local Similarity 86.0%; Pred. No. 2.1e-221;

Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

QY 1 MSISDEVNFLVRYLQESGFHSAFTFGIESHISOSNINGALVPPAALISIIQKLOVY 60

DB 52 MSITSDEVNFLVRYLQESGFHSAFTFGIESHISOSNINGTLVPPAALISIIQKLOVY 111

QY 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVVQTQAYVDKLAQHAAAAAATAAT-- 118

DB 112 EAEISINEDGTVFGRPIESLSLIDAVMPDVVQTQAFREKLAQQAQAAAAATAA 171

QY 119 -----NOGSAGKNGENTANGENGATIANHTDMMEVDGDIIPSNKAVVLRG 167

DB 172 ARAATTTSAGVSHQNPSPKREATVNGEENRAHSV--NNHAKPWEIDGEVPIPSKATVLRG 230

QY 168 HESEVFICAWNPSVLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGQDVPNSKD 227

DB 231 HESEVFICAWNPSVLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGQDVPNSKD 290

QY 228 VTSLDWNSGTLTATGSDYDGFARIMTKDGNLASTLGHKGPIFALKWKKGNFILLSAGVD 287

DB 291 VTSLDWNTNGTLTATGSDYDGFARIMTKDGNLASTLGHKGPIFALKWKKGNFILLSAGVD 350

QY 288 KTTIIMDAHTGEAKQOFPFHSAAPALVDWQSNNTFASCTDMCIHVCKLQGDREIKTFQG 347

DB 351 KTTIIMDAHTGEAKQOFPFHSAAPALVDWQSNNTFASCTDMCIHVCKLQGDREIKTFQG 410

QY 348 HTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKEIYTIKWSPTGPTN 407

DB 411 HTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKEIYTIKWSPTGPTN 470

QY 408 NPNANILMASAGFSDTVRLWDVDRGICHTLTKEHPEVYSVAFSPDGRYLASGSDKCVH 467

DB 471 NPNANILMASAGFSDTVRLWDVDRGICHTLTKEHPEVYSVAFSPDGRYLASGSDKCVH 530

QY 468 IWNQTGALVHSYRGTGGIFVVCWNAAGDKVGSASDGSVCVLDLRK 514

DB 531 IWNQTGALVHSYRGTGGIFVVCWNAAGDKVGSASDGSVCVLDLRK 577

RESULT 4

ABP41760

ID ABP41760 standard; protein; 542 AA.

XX AC ABP41760;

DT 22-AUG-2002 (first entry)

XX Human ovarian antigen HNOX38, SEQ ID NO:2892.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;

KW Human cancer; breast cancer; tumour; reproductive system disorder;

KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 OS Homo sapiens.
 XX WO200200677-A1.
 PN 03-JAN-2002.
 PD 07-JUN-2001; 2001WO-US018569.
 PF 07-JUN-2000; 2000US-0209467P.
 PR (HUMA-) HUMAN GENOME SCI INC.
 XX Birse CE, Rosen CA;
 PI WPI; 2002-147878/19.
 DR N-PSDB; ABQ54837.

PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 PT cancer), immune disorders, cardiovascular disorders and neurological
 PT diseases.

PS Claim 11; SEQ ID NO 2892; 2922pp; English.

CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 542 AA;

Query Match 85.8%; Score 2342; DB 5; Length 542;
 Best Local Similarity 85.4%; Pred. No. 2.9e-212;
 Matches 434; Conservative 32; Mismatches 28; Indels 14; Gaps 2;

QY 20 GFSHSAFTGIESHSQNSINGALVPPAALISITOKGLOYVEAEVSEINEDGTLFEGRP 79
 DB 36 GFSHSAFTGIESHSQNSINGALVPPAALISITOKGLOYVEAEVSEINEDGTLFEGRP 95

QY 80 SLSLIDAVPDVVTQQAAYRKLACQAAHAAAAAARAT-----NQGSAGN 126
 DB 96 SLSLIDAVPDVVTQQAAYRKLACQAAHAAAAAARAT-----NQGSAGN 155
 QY 127 GENTANGENGARTIANNHTDMWEVDGVEIPSNKAVLRLGHESEVFFICAMNPVSDLLVS 186
 DB 156 REATYNGEENRAHSV-NNHAKPWEIDGVEIIPSSKATVLRGHESEVFFICAMNPVSDLLAS 214
 QY 187 GSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPNSKOVTSLDWNSEGTLLATGSDYD 246
 DB 215 GSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPNSKOVTSLDWNSEGTLLATGSDYD 274
 QY 247 GFARIWTKDGNLSTLQGHKGPFIKWNKNGNFILSAGVDKTTIWDATGEAKQQPFF 306
 DB 275 GFARIWTEGNGLSTLQGHKGPFIKWNKNGNYILSAGVDKTTIWDATGEAKQQPFF 334
 QY 307 HSAPALDVQSNNTFASCSSTDMCIHVCKLGGDRPIKTFQHTNEVNAIKWDPTGNLLAS 366
 DB 335 HSAPALDVQSNNTFASCSSTDMCIHVCKLGGDRPIKTFQHTNEVNAIKWDPTGNLLAS 394
 QY 367 CSDDMTLKIWSMKQDNCVHDLOAHNKEIYTIKWSGTGCTNNPNANLALASPDSTVRL 426
 DB 395 CSDDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSGTGCTNNPNANLALASPDSTVRL 454
 QY 427 WDVDRGICITHTLTKHQEPVYVAFSPDGRYLAGSFDKCVKHIWNTQTGALVHSYRGCGI 486
 DB 455 WDIERGVCVTHLTTKHQEPVYVAFSPDGRYLAGSFDKCVKHIWNTQSGNLVHSYRGCGI 514
 QY 487 FEVCWNAAGDKVGASASDGSVCVLDLRK 514
 DB 515 FEVCWNAAGDKVGASASDGSVCVLDLRK 542

RESULT 5
 ABP51424
 ID ABP51424 standard; protein; 395 AA.
 XX
 AC ABP51424;
 XX
 DT 03-SEP-2002 (first entry)
 XX
 DE Human MDDT SEQ ID NO 446.
 XX
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatocytic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;
 KW antiallergic; antianemic; antiasthmatic; antiatherosclerotic; antigout;
 KW neuroprotective; antirheumatic; antiarthritic.
 OS Homo sapiens.
 XX WO2002040715-A2.
 XX 23-MAY-2002.
 PD 06-SEP-2001; 2001WO-US027628.
 PF 05-SEP-2000; 2000US-0229747P.
 PR 05-SEP-2000; 2000US-0229748P.
 PR 05-SEP-2000; 2000US-0229749P.
 PR 05-SEP-2000; 2000US-0229750P.
 PR 05-SEP-2000; 2000US-0229751P.
 PR 05-SEP-2000; 2000US-0230583P.
 PR 06-SEP-2000; 2000US-0230585P.
 PR 06-SEP-2000; 2000US-0230514P.
 PR 06-SEP-2000; 2000US-0230515P.
 PR 06-SEP-2000; 2000US-0230517P.
 PR 06-SEP-2000; 2000US-0230518P.
 PR 06-SEP-2000; 2000US-0230519P.
 PR 06-SEP-2000; 2000US-0230595P.
 PR 06-SEP-2000; 2000US-0230597P.

PR 06-SEP-2000; 2000US-0230598P.
 PR 06-SEP-2000; 2000US-0230599P.
 PR 06-SEP-2000; 2000US-0230610P.
 PR 06-SEP-2000; 2000US-0230612P.
 PR 06-SEP-2000; 2000US-0230855P.
 PR 06-SEP-2000; 2000US-0230988P.
 PR 06-SEP-2000; 2000US-0230989P.
 PR 07-SEP-2000; 2000US-0230951P.
 PR 07-SEP-2000; 2000US-0231163P.
 PR 07-SEP-2000; 2000US-0231167P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;
 XX Hillman JR, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE;
 XX Dahl CR, Moniyama MG, Bradley DL, Rohatgi SD, Harris B;
 XX Roseberry AM, Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V;
 XX Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
 DR WPI; 2002-527544/56.
 DR N-PSDB; ABQ72641.
 XX Novel human disease detection and treatment polypeptide, useful in
 XX diagnosis, prevention or treatment of cell proliferative disorders e.g.
 XX arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.
 XX AIDS.
 XX Claim 14; Page 578; 619pp; English.
 XX The invention relates to an isolated human disease detection and
 XX treatment (MDDT) polypeptide (I) selected from a polypeptide having a
 XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the
 XX specification, a naturally occurring polypeptide comprising a sequence
 XX having at least 90% identity to (I) or a biologically active or
 XX immunogenic fragment of (I). (I) is useful for screening a compound for
 XX effectiveness as an agonist or antagonist, for screening a compound that
 XX specifically binds (I) or modulates the activity of (I), and for
 XX preparing a polyclonal or monoclonal antibody by hybridoma technology.
 XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 XX screening a compound for effectiveness in altering expression of a target
 XX polynucleotide comprising. Oligonucleotides and antibodies are useful for
 XX detecting MDDT in a sample or for assessing toxicity of a test compound,
 XX in a diagnostic test for a condition or a disease associated with the
 XX expression of MDDT in a biological sample, for detecting (I) in a sample,
 XX and for purifying (I) from a sample. A composition comprising (I), an
 XX agonist or antagonist is useful for treating a disease or condition
 XX associated with decreased or increased expression of functional MDDT. (I)
 XX or (II) are useful for diagnosing, treating or preventing disorders
 XX associated with aberrant expression of MDDT, where the disorders are
 XX selected from a cell proliferative disorder such as arteriosclerosis,
 XX cirrhosis, hepatitis, psoriasis, and cancer and an
 XX autoimmune/inflammatory disorder such as AIDS. Addison's disease,
 XX allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
 XX rheumatoid arthritis. (II) are useful for creating knockin humanised
 XX animals or transgenic animals to model human diseases, in somatic or
 XX germ-line gene therapy, to generate a transcript image of a tissue or cell
 XX type, for detecting differences in the chromosomal location due to
 XX translocation or inversion among normal, carrier or affected individuals
 XX and as hybridisation probes for mapping naturally occurring genomic
 XX sequences
 XX Sequence 395 AA;
 XX
 XX Query Match 74.2%; Score 2027; DB 5; Length 395;
 XX Best Local Similarity 98.5%; Pred. No. 1.1e-182;
 XX Matches 385; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 Qy 1 MSISDEVNFIYRLQESGFSHGAFTGIESHTSQSNINGALVPPAALISITKGLQYV 60
 Db 5 MSISDEVNFIYRLQESGFSHGAFTGIESHTSQSNINGALVPPAALISITKGLQYV 64
 Qy 61 EAEVSIINEDGTLFGRPTIESLSLIDAVMPVQTRQAYRDKLAQOCHAAAAAATNQ 120
 Db 65 EAEVSIINEDGTLFGRPTIESLSLIDAVMPVQTRQAYRDKLAQOCHAAAAAATNQ 124

121 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPSKAVVLRGHESEVFICAWNPV 180
 125 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFICAWNPV 184
 181 SLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSPNKDVTSLDMNSEGTL 240
 185 SLLASGSGDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSPNKDVTSLDMNSEGTL 244
 241 ATGSDGPFARIWTKDGNLASTIGQHKGPFAKWNKKNFILSAGVDKTTIIDAHTGEA 300
 245 ATGSDGPFARIWTKDGNLASTIGQHKGPFAKWNKKNFILSAGVDKTTIIDAHTGEA 304
 301 KQOPFPHSAPALDWDQSNNTFASGSDTMCIIHVCKLQDRIKTFQGHNEVNAIKWDPT 360
 305 KQOPFPHSAPALDWDQSNNTFASGSDTMCIIHVCKLQDRIKTFQGHNEVNAIKWDPT 364
 361 GNLASCSDDMTLKWSMKQDNCVHDLOAHN 391
 365 GNLASCSDDMTLKWSMKQDNCVHDLOAHN 395

RESULT 6
 ABB60376
 ID ABB60376 standard; protein; 700 AA.
 XX
 XX ABB60376;
 XX AC
 XX AC
 XX AC
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 7920.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 XX (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656960/75.
 DR N-PSDB; ABL04479.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signaling and cell-cell
 XX interactions.
 XX
 PS Disclosure; SEQ ID NO 7920; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signalling and
 XX cell-cell interactions in higher eukaryotes for the development of
 XX insecticides, therapeutics and pharmaceutical drugs. The invention
 XX discloses genomic DNA sequences (ABL16175-ABL30511), expressed DNA
 XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS5737-
 XX ABBS7072). The sequence data for this patent did not form part of the
 XX printed specification, but was obtained in electronic format directly
 XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 700 AA;
 XX
 XX Query Match 69.8%; Score 1906.5; DB 4; Length 700;

Best Local Similarity 54.8%; Pred. No. 7.2e-171; Matches 385; Conservative 49; Mismatches 77; Indels 191; Gaps 8;

QY 1 MSISDEVNLFVRYLQESGSHAFTEGIESHSQSNGALVPPAALISIIQKGLQYV 60
DB 1 MSFSSDEVNLFVRYLQESGSHAFTEGIESHSQSNGALVPPAALISIIQKGLYT 60

QY 61 EAEVSIINDEGTLFQGRPTIESLIDAVMPDV-----VQTRQ----- 96
DB 61 EAEVSIINDEGTLFQGRPTIESLIDAVMPDV-----VQTRQ----- 96

QY 97 -----QAYRDKLA-----QQHAAAAAATAATNOGSAKNGENT 130
DB 97 -----QAYRDKLA-----QQHAAAAAATAATNOGSAKNGENT 130

QY 119 NAKPEIKIEPTGVAGSAGGNKIAGSTTGTPTDQASAEVDSGNAAGGTAGNG 178
DB 119 NAKPEIKIEPTGVAGSAGGNKIAGSTTGTPTDQASAEVDSGNAAGGTAGNG 178

QY 131 ANGER----- 135
DB 131 ANGER----- 135

QY 179 AGGNOASTGGNSTSTPAGGDLAAGSOKSONSEAGSSSGNAGNANATSTDDAASS 238
DB 179 AGGNOASTGGNSTSTPAGGDLAAGSOKSONSEAGSSSGNAGNANATSTDDAASS 238

QY 136 ---NG-----AHTIANNTDM----- 148
DB 136 ---NG-----AHTIANNTDM----- 148

QY 239 TSTNGNSTSSVEQPTSGLTGAGTGTSTPDAASGASATGSKAPSAVIRVGAQ 298
DB 239 TSTNGNSTSSVEQPTSGLTGAGTGTSTPDAASGASATGSKAPSAVIRVGAQ 298

QY 149 -----MEVDGDIPIPSKAVVLRGHESEV 172
DB 149 -----MEVDGDIPIPSKAVVLRGHESEV 172

QY 299 GNVQSGSSNAQSSAPSGTISSTSGGAGTPEALVPMIDENIBIPESKARVLRGHESEV 358
DB 299 GNVQSGSSNAQSSAPSGTISSTSGGAGTPEALVPMIDENIBIPESKARVLRGHESEV 358

QY 173 FICANPVSDLLVSGSGSTARIWNLSENSTSGPTQVLRHCIEGGDVPSPKNDVTSLD 232
DB 173 FICANPVSDLLVSGSGSTARIWNLSENSTSGPTQVLRHCIEGGDVPSPKNDVTSLD 232

QY 359 FICANPVSDLLVSGSGSTARIWNLSENSTSGPTQVLRHCIEGGDVPSPKNDVTSLD 417
DB 359 FICANPVSDLLVSGSGSTARIWNLSENSTSGPTQVLRHCIEGGDVPSPKNDVTSLD 417

QY 233 WNSEGTILATGSDYGFARITWTKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTII 292
DB 233 WNSEGTILATGSDYGFARITWTKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTII 292

QY 418 WNCDSGLATGSDYGFARITWTKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTII 477
DB 418 WNCDSGLATGSDYGFARITWTKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTII 477

QY 293 WDAHTGEAKQPPHSAFALVDWQSNVTFASCSTDMCIHVCKLGQDRPIKTFQGHNEV 352
DB 293 WDAHTGEAKQPPHSAFALVDWQSNVTFASCSTDMCIHVCKLGQDRPIKTFQGHNEV 352

QY 478 WDAHTGEAKQPPHSAFALVDWQSNVTFASCSTDMCIHVCKLGQDRPIKTFQGHNEV 537
DB 478 WDAHTGEAKQPPHSAFALVDWQSNVTFASCSTDMCIHVCKLGQDRPIKTFQGHNEV 537

QY 353 NAIKWDPTGNLLASCSDDMTLKWSMKDNCVHDLQAHNKEIYTIKWSPTGPGNNPNAN 412
DB 353 NAIKWDPTGNLLASCSDDMTLKWSMKDNCVHDLQAHNKEIYTIKWSPTGPGNNPNAN 412

QY 538 NAIKWDPTGNLLASCSDDMTLKWSMKDNCVHDLQAHNKEIYTIKWSPTGPGNNPNAN 597
DB 538 NAIKWDPTGNLLASCSDDMTLKWSMKDNCVHDLQAHNKEIYTIKWSPTGPGNNPNAN 597

QY 413 LMLASAFDSTVRLWVDRGCIHTLTKHQBPFVSVAFSPDGRVLASGDFKCVHIWNTQ 472
DB 413 LMLASAFDSTVRLWVDRGCIHTLTKHQBPFVSVAFSPDGRVLASGDFKCVHIWNTQ 472

QY 598 LILASAFDSTVRLWVDRGCIHTLTKHQBPFVSVAFSPDGRVLASGDFKCVHIWNTQ 657
DB 598 LILASAFDSTVRLWVDRGCIHTLTKHQBPFVSVAFSPDGRVLASGDFKCVHIWNTQ 657

QY 473 TGAHVSYRGTGIGFVWCNNAAGDVAGASADSGVVCVLDLAK 514
DB 473 TGAHVSYRGTGIGFVWCNNAAGDVAGASADSGVVCVLDLAK 514

QY 658 TGAHVSYRGTGIGFVWCNNAAGDVAGASADSGVVCVLDLAK 699
DB 658 TGAHVSYRGTGIGFVWCNNAAGDVAGASADSGVVCVLDLAK 699

RESULT 7
ID ABG21351
XX ABG21351 standard; protein; 584 AA.
AC ABG21351;
XX
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #21342.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
ED
XX 30-MAR-2001; 2001WO-US008631.
PF
XX
XX

PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Dmanac RT, Liu C, Tang YT;
XX
XX MPI: 2001-639362/73.
XX N-PSDB; AAS85538.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 51710; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 584 AA;
Query Match 69.3%; Score 1891.5; DB 4; Length 584;
Best Local Similarity 64.6%; Pred. No. 1.4e-169;
Matches 368; Conservative 34; Mismatches 39; Indels 129; Gaps 4;
QY 16 LQESGSHAFTEGIESHSQSNGALVPPAALISIIQKGLQYVEAEVSIINDEGTLFDG 75
DB 16 LQESGSHAFTEGIESHSQSNGALVPPAALISIIQKGLQYVEAEVSIINDEGTLFDG 75

QY 62 LEVKGFSHAFTEGIESHSQSNGALVPPAALISIIQKGLQYVEAEVSIINDEGTLFDG 121
DB 62 LEVKGFSHAFTEGIESHSQSNGALVPPAALISIIQKGLQYVEAEVSIINDEGTLFDG 121

QY 76 RPTESLSLIDAVMPDVQTRQAYRDKLAQOHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 124
DB 76 RPTESLSLIDAVMPDVQTRQAYRDKLAQOHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 124

QY 122 RPTESLSLIDAVMPDVQTRQAYRDKLAQOHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 181
DB 122 RPTESLSLIDAVMPDVQTRQAYRDKLAQOHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 181

QY 125 KRGENTANGEANGAHTIANNHTDMMEVDGDIPIPSKAVVLRGHESEVFIKAWNPVSDLL 184
DB 125 KRGENTANGEANGAHTIANNHTDMMEVDGDIPIPSKAVVLRGHESEVFIKAWNPVSDLL 184

QY 182 KUREATVNGEANGAHEI-NNHKSPMEIDGDIPIPSKAVVLRGHESEVFIKAWNPVSDLL 240
DB 182 KUREATVNGEANGAHEI-NNHKSPMEIDGDIPIPSKAVVLRGHESEVFIKAWNPVSDLL 240

QY 185 VSGSGDSTARIWNLSENSTSGPTQVLRHCIEGGDVPSPKNDVTSLDWNSGTLATGS 244
DB 185 VSGSGDSTARIWNLSENSTSGPTQVLRHCIEGGDVPSPKNDVTSLDWNSGTLATGS 244

QY 241 ASGSGDSTARIWNLSENSTSGPTQVLRHCIEGGDVPSPKNDVTSLDWNSGTLATGS 300
DB 241 ASGSGDSTARIWNLSENSTSGPTQVLRHCIEGGDVPSPKNDVTSLDWNSGTLATGS 300

QY 245 YDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTIIWDAHTGEAKQOF 304
DB 245 YDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTIIWDAHTGEAKQOF 304

QY 301 YDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTIIWDAHTGEAKQOF 310
DB 301 YDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTIIWDAHTGEAKQOF 310

QY 305 PHSAPALVDWQSNVTFASCSTDMCIHVCKLGQDRPIKTFQGHNEV 349
DB 305 PHSAPALVDWQSNVTFASCSTDMCIHVCKLGQDRPIKTFQGHNEV 349

QY 311 ---NAPALVDWQSNVTFASCSTDMCIHVCKLGQDRPIKTFQGHNEV 367
DB 311 ---NAPALVDWQSNVTFASCSTDMCIHVCKLGQDRPIKTFQGHNEV 367

QY 350 -----NEYNAIKWDPT 360
DB 350 -----NEYNAIKWDPT 360

QY 368 KLTMTTEGRLRPKTFCSGSGALLPPAGRRPHLLTGPDIFFKRLSALFQNEVNAIKWDPS 427
DB 368 KLTMTTEGRLRPKTFCSGSGALLPPAGRRPHLLTGPDIFFKRLSALFQNEVNAIKWDPS 427

```
QY 361 GNLLASCSDDMTLKINSMKQDNCVHDLQAHNKSIYTIKWSPTGPTGNTNPNANLMLASAF 420
Db 428 GMLLASCSDDMTLKINSMKQDNCVHDLQAHNKSIYTIKWSPTGPTGNTNPNANLMLASAF 487
QY 421 DSTVRLNDVDRGICHTLTKEHBPVYVAFSPDGRYLASGSPKCVHIMNTQTGALVHSY 480
Db 488 DSTVRLNDVDRGICHTLTKEHBPVYVAFSPDGRYLASGSPKCVHIMNTQTGALVHSY 547
QY 481 RGTGGIFEVCMNAGDKVGASASDGSVCVL 510
Db 548 QGTGGIFEVCMNAGDKVGASASDGSVCIL 577

RESULT 8
ID AAO04385 standard; protein; 208 AA.
XX AC AAO04385;
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 18277.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US004927.
XX PR 28-FEB-2000; 2000US-00515126.
XX PR 18-MAY-2000; 2000US-00577409.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX DR N-PSDB; AAI84316.
XX PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX PT and treating e.g. leukemia, inflammation and immune disorders.
XX PS Claim 20; SEQ ID NO 18277; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation. Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 208 AA;
Query Match 41.8%; Score 1142; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 2.4e-99;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 213 HCIREGGQDVPFNKDVTSLDWNSGTLATGSGYDGFARIWTKDGNLASTLQGHKGPFFAL 272
Db 273 KWKKGKGNFTLSAGVDKTTIWDHAHTGEAKQOPPHSAFALDWDQSNNTFASCSDDMCIH 332
61 KWKKGKGNFTLSAGVDKTTIWDHAHTGEAKQOPPHSAFALDWDQSNNTFASCSDDMCIH 120
333 VCKLQGDPRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKINSMKQDNCVHDLQAHNK 392
121 VCKLQGDPRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKINSMKQDNCVHDLQAHNK 180
393 EYTIKWSPTEGPTGNTNPNANLMLASAS 419
181 EYTIKWSPTEGPTGNTNPNANLMLASAS 207

RESULT 9
ID ABG21350 standard; protein; 167 AA.
XX AC ABG21350;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #21341.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS85337.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 51709; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have application in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
```

```
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 167 AA;
Query Match 30.2%; Score 824; DB 4; Length 167;
Best Local Similarity 86.8%; Pred. No. 2,1e-69;
Matches 145; Conservative 13; Mismatches 9; Indels 0; Gaps 0;
QY 250 RIWTKDGNLSTLGHKGPFPALKNKGNFPLSAGVDKTTIWDHTGEAKQFPFHS 309
Db 1 RIWTEGNLAITLCOHKGPIFALKWKNKGNVLSAGVDKTTIWDHTGEAKQFPFHS 60
QY 310 PALDWDQNSNTFASCTDMCIHVCKLGDRPIKTFQGHTEVNAIKWDPTGNLLASCSD 369
Db 61 PALDWDQNSNTFASCTDMCIHVCKLGDRPIKTFQGHTEVNAIKWDPTGNLLASCSD 120
QY 370 DMTLKWSMKQNCVHDLQAHNKETIYTKWSPGPTGNTNPNANMLA 416
Db 121 DMTLKWSMKQNCVHDLQAHNKETIYTKWSPGPTGNTNPNANMLA 167
RESULT 10
ABRS52980
ID ABR52980 standard; protein; 535 AA.
AC ABR52980;
DT 20-JUN-2003 (first entry)
XX Protein sequence #SEQ ID 825.
DE Multi-protein complex; eukaryote; drug target; diagnosis.
KW Saccharomyces cerevisiae.
OS Saccharomyces cerevisiae.
XX EP1258494-A1.
XX 20-NOV-2002.
XX 20-DEC-2001; 2001EP-00130253.
XX 15-MAY-2001; 2001EP-00111774.
XX (CELL-) CELLZONE AG.
XX Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
XX Marzioch M, Schultz JD, Superti-Furga GD;
XX WPI: 2003-250078/25.
XX N-PSDB; ACC61022.
XX New isolated protein complexes useful for diagnosing a disease or
XX disorder, or as a target for an active agent of a pharmaceutical,
XX preferably a drug target in the treatment or prevention of disease or
XX disorder.
XX Disclosure; SEQ ID NO 825; 17pp + Sequence Listing; English.
XX The invention relates to multiprotein complexes from eukaryotes. Proteins
XX of the invention and DNA sequences encoding them are given in records
XX ABR52568-ABRS3903 and ACC60610-ACC61944 respectively. The complexes are
XX obtainable by using a protein as a bait and isolating the set of proteins
XX which is attached thereto from cells. Such protein complexes may comprise
XX up to 30 distinct proteins. Protein complexes of the invention are useful
XX for diagnosing a disease or disorder, or as a target for an active agent
XX of a pharmaceutical, preferably a drug target in the treatment or
XX prevention of a disease or disorder. Note: The sequence data for this
XX patent is not represented in the printed specification, but is based on
XX sequence information supplied by the European Patent Office. The complete
XX document is available on CD-ROM
XX Sequence 535 AA;
```

```
Query Match 18.3%; Score 498.5; DB 6; Length 535;
Best Local Similarity 25.4%; Pred. No. 8.7e-38;
Matches 143; Conservative 101; Mismatches 239; Indels 79; Gaps 16;
QY 1 MTSISDEVNLFVRYLQESGSHSAFTFGIESHISQSNIN-CALVPPAALISIIQKQY 59
Db 1 MTSISDEVNLFVRYLQESGSHSAFTFGIESHISQSNIN-CALVPPAALISIIQKQY 60
QY 60 VRAEVSINEDG--TLFDRPIESLSLDVNPDPVVTQRAYRQKLAQCHAAAAA 116
Db 61 TESELMVDKSGDISALNEHLSDFNLVQALQID-----KEKFFE--ISSEGREFTL 109
QY 117 ATNCOGSAGKNGENTANGEANGAHTIANHHTDMW-EVDGDVEIPSNKAVLRGHESEVFI 175
Db 110 ETNSE-SNKAEGDASTVERETQEDDTNSIDSSDDLDGFVKI--LKEIV--KLDNIYSS 163
QY 176 ANPYSD-LLVSGSGDSTARIWNLSNSTSGPTQ-----LVRHCIREGGODVPSNKDV 228
Db 164 TWPPLDESILAYCEKNSVARLARIIVETDQEGKKYWKLTIIAELRHPFALSASSGKTTNQV 223
QY 229 TSLDWNSECTLLATSGYDGFARIWTKDGNLSTLGHKGPFPALKNKGNFPLSAGVDK 288
Db 224 TCLAWSHDGNSIVTGVENGELRLNKTGALLNLVNFHRAPIVSVKWKDGTIIISNDVEN 283
QY 289 TTIWDAHTGEAKQFPF-----HSAP---ALDWDQNSNTFASCTDMCIHV 333
Db 284 VTILMNVISGTVMQHPFELKGTGSSINAENHSGDGLGVDEWVDDDKFVIPGPKGAIFV 343
QY 334 CKLGQDRPIKTFQGHTEVNAIKWDPTGNLLASCDDMTLKWSMKQNCVHDLQAHNK 393
Db 344 YQITEKTPTKLIGHHGPISVLEFNDTKLLSASDDGTLRWGGNGNSQNCFFYGHQS 403
QY 394 IYTIKWSPTGPTNPNANMLASASFDSTVALWDVDRGICITHTLTKHQEPVYSVAFSPD 453
Db 404 IVSASWV-----GDDKVISCSMDGSLVLSLQNTLLALSIVDGVPFAGRIQSD 453
QY 454 GRYLASGSPDKVHIW-----NTQTGAL-----VHSYRGTTGIFVC 490
Db 454 GQKYAVAFMDGQVNVYDLKLNKSKRSRYGNRDLNPLPLIPLIYASQSDNDYIFDLS 513
QY 491 WNAAGDKVGA--SASDGSVCVL 510
Db 514 WNCAGNKISVAYSLQEGSVVAI 535
RESULT 11
ABRS3774
ID ABR53774 standard; protein; 515 AA.
AC ABR53774;
XX 20-JUN-2003 (first entry)
XX Protein sequence #SEQ ID 2413.
XX Multi-protein complex; eukaryote; drug target; diagnosis.
XX Saccharomyces cerevisiae.
XX EP1258494-A1.
XX 20-NOV-2002.
XX 20-DEC-2001; 2001EP-00130253.
XX 15-MAY-2001; 2001EP-00111774.
XX (CELL-) CELLZONE AG.
XX Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
XX Marzioch M, Schultz JD, Superti-Furga GD;
XX WPI: 2003-250078/25.
```


Query Match 13.7%; Score 374.5; DB 6; Length 414;
 Best Local Similarity 26.7%; Pred. No. 3.2e-26;
 Matches 111; Conservative 55; Mismatches 165; Indels 85; Gaps 13;

QY 83 LIDAYMPVQTRQAYRDKLAQQ-----HAAAAAATAAQQGSA-KNGE-- 128
 DB 54 LITASRTQVKLLIQRLQKLGQSNHTFYLKVLKAHILPLTNVALNKSGSCFITGSDY 113
 QY 129 -----NTANGENGCAHTIANNHTDMEVDGVDVIPSNAVVLGHESEVEFICAW-NPVS 181
 DB 114 RTCKLWDITASGEBLN-----TLGHRNVVYTAIFNPNPYG 147

QY 162 DLLVSGSDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLLA 241
 DB 148 DKATGSDFDKCKLWSVETG-----CYHTR-----GHTAEIVCLSFNPQSTLVA 193

QY 242 TGSYDGFARIWTKOGLASTIGQHKGPITFALKWKKGNFILSAGVDKTTIWDHTGEAK 301
 DB 194 TGSMDTTAKLDIQNGEELTLRGHSABEISLSFNTSGDRIITGSDFTVVVWMDADTGRV 253

QY 302 QQFPFSA--PALVDVQSNNTFASCS---TDMCHVCKL---GQDRPIKTFQGTNEVN 353
 DB 254 NILIGHCAEISSAFNW-----DCSLILTGSDMKTKCLWDATNGKCVALTGDDDEIL 306

QY 354 AIKWDPGNTLLASCDMDTLKIWSMKDQNCVHDLOAHNKEIYTIKWSPGTGTPNPNANL 413
 DB 307 DSCFDYTGKLIATASADGTAFISAATRKCIKLEGEHEGEEKISF-----NPQGNH 358

QY 414 MLASFSFTVRLWDVDRGICHTLTKEHPVYSVAESPDRYLASGDFKCVHIW 469
 DB 359 LLTGSS-DKTAIRIWDQOTGQCLQVLEHTDEIFSCAFNYKGNIVITGSKONTCKRIW 413

RESULT 13

AA85881
 ID AAR85881 standard; protein; 514 AA.
 XX AAR85881;
 AC AAR85881;
 XX
 DT 13-SEP-1996 (first entry)
 DE WD-40 domain-contg. YCW2 protein.
 XX
 KW WD40 repeat region; beta-transducin; protein-protein interaction; drug;
 KW intracellular signalling; protein kinase C; homology; motif; modulator;
 KW receptors of activated protein kinase; enzyme activity; isozyme; human.
 XX
 OS Synthetic.
 XX
 FN WO9521252-A2.
 XX
 PD 10-AUG-1995.
 XX
 PF 31-JAN-1995; 95WO-US001210.
 XX
 PR 01-FEB-1994; 94US-00190802.
 XX
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 XX
 PI Mochly-Rosen D, Ron D;
 XX
 DR WPI; 1995-283772/37.
 XX
 PT New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the
 PT activity of a protein, eg. protein kinase C, which interacts with a
 PT protein contg. a WD-40 region.
 XX
 PS Example 5; Page 168-170; 351pp; English.
 XX
 CC Proteins AAR8581-92 are protein which contain at least one WD-40 (also
 CC called beta-transducing homologous) amino acid repeat motifs. The WD-40
 CC regions are involved in protein-protein interactions between proteins

CC involved in intracellular signalling. An example of such an interaction
 CC is between protein kinase C and receptors of activated protein kinase
 CC (RACK), esp. RACK-1 (AAR85850). Proteins AAR8581-82 were isolated based
 CC on homology with beta-transducin, whereas proteins AAR85882-92 were
 CC isolated based on homology with the WD-40 consensus sequence (AAR85893).
 CC The proteins were used to construct the peptides AAR84928-R85063 and
 CC AAR85786-R85842. The peptides can be used to identify target proteins
 CC contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of
 CC proteins involved in protein-protein interaction and to screen for drugs
 CC that will affect protein-protein interaction involving WD-40 domains
 XX
 XX Sequence 514 AA;

Query Match 13.7%; Score 374.5; DB 2; Length 514;
 Best Local Similarity 23.6%; Pred. No. 4.5e-26;
 Matches 125; Conservative 85; Mismatches 167; Indels 153; Gaps 21;

QY 38 NINGALVPPAALISITQKLOVVEAEVINEGDTL-----PDGR-----PIESLSLIDA 86
 DB 44 NVGGALRVPCA---ISEKLEELNLQNGTSDDPVYTFSTCTIQGKASDPVKTTIDITDN 100
 QY 87 VMPDVVQTRQAYRDKLAQOHA 146
 DB 101 LVSSLIKPGYNSTEQIITLLYTPRAVFKVPTVRSSA----- 138

QY 147 DMMEVDGVEIIPSNKAVLVRGHESEVEFICAMNP-VSDLLVSGSGDSTARIWNLSNSTSG 205
 DB 139 -----IAGHGSTILCSAFAPHTSSRMVTGAGDNTARIWDC---DTQT 177

QY 206 PTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLLATGSDGFARIW-TKQGN-LASTLG 263
 DB 178 PHMTLKGH-----YMWLVCSWSPDGEVATGSMNTILNDPKSGOCLGDALR 226

QY 264 QHKGPITFALKWN-----KQGN--FILSAGVDKTTIWDHTGEAKQPPFHSAPALDVM 316
 DB 227 GHSKWTISLWEPHILVKPGRPLASSSKOGTIKINDTVS----- 267

QY 317 QSNNTFASCSGDMCHIVCKLGQDRPIKTFQGTNEVNALKWDPTGNLLASCDMDTLKIWI 376
 DB 268 -----RVCCY-----TWSGHTNSVSCVKWGGQG-LLYSGSHDRTVRVW 304

QY 377 SM-KQDNCVHDLQAHNKEI-----YTIK---WSPTG---PGT----- 406
 DB 305 DINSQGRGINILKSHAHVWNLSTLDYALRIGAFDHTGKXKPEEAKKALENYEKIC 364

QY 407 -NNPNANILMASASDSTVRLWDVDRGI-CIHTLTKEHPVYSVAESPDRYLASGSDK 464
 DB 365 KKGNSSEEMVMTASDDYTMFLNPLKSTKPIARMTGHQKLVNHHVAFSPDGRIVYSASFDN 424

QY 465 CVHIWNTQTGALVHSGYRG-TGGIFEVCAAGDKVGSASDGSVCVLDLR 513
 DB 425 SIXLWDGRDGRKEISTFRGHIAVSYQVAMSSDCELLVSCSKDTTLKVDVR 474

RESULT 14

AB859486
 ID ABB59486 standard; protein; 481 AA.
 XX
 XX ABB59486;
 AC ABB59486;
 XX
 DT 26-MAR-2002 (first entry)
 DE
 DE Drosophila melanogaster polypeptide SEQ ID NO 5250.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX

```
PF 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL03589.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 5250; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signaling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 481 AA;
XX
XX Query Match 13.4%; Score 366.5; DB 4; Length 481;
XX Best Local Similarity 27.5%; Pred. No. 2.3e-25;
XX Matches 128; Conservative 61; Mismatches 175; Indels 101; Gaps 21;
XX
XX QY 115 AATNQ-----GSAKNGENTAN-----GRENGAHTIANNHTDMEVDG--VEIPSNKA 162
XX Db 35 AGITQQLGLIGNALLKNEEATPYLFVGEDEIKKSLEDT-LDLASVDITENVIDIYVQPQ 93
XX
XX QY 163 VLVR-----CHESEVFICAWNPVSDLLVSGSGDSTARIWNLSN-----STSGP 206
XX Db 94 AVFKVRPVTRCTSMPEGAAVSLNPSPDGAHLASGSGTTLVRLDNLNTETPHFTCTGH 153
XX
XX QY 207 TQLVLRHCIR-----EGQD-----VPSNKDVTSLDW-----N 234
XX Db 154 KQWVL--CVSWAPDKRLASGCKAGSI IWDPTGQKGRPLSGHKKHINCLAWEPYHRD 211
XX
XX QY 235 SEGTLATGSDGPARTW-TKDGNIASLTGQHKGPIFALKWKKGNFILSAGVDKTTI 293
XX Db 212 PECKLASAGDGCRIWDVLGQCLMNIAGHTNAVTAVRWGAG-LIYTSSKDRIVKMW 270
XX
XX QY 294 DAHTGEAKQFPFPHSPALPDVDMQSNNTFASCTDMCIHVCKLQDPRPIK-TFQGH---- 348
XX Db 271 RAADGILCRTFSGHA-----HWNN--TALSTD--YVLRTPGFPHVDRSKSHLSLS 318
XX
XX QY 349 TNEV-----NAIKWDPGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWS 400
XX Db 319 TEELQESALKRYQAVCPDEVESLV-SCSDNTLYLWRNNQKCVERTMGTQNVNDVKYS 377
XX
XX QY 401 PTGPGTNNPNNANMLASGSPSTVRLMDVDRGICHTLTKHQBEPVYSVAFSPDGRYLASG 460
XX Db 378 PDVK-----LIASGPDKSVLRLRASDGGYMATFRGHVQAVYTVANSADSLIVSG 428
XX
XX QY 461 SFDKCVHNTQTGALVHSYRG-TGGIFEVCAWNAAGDKVVASAD 504
XX Db 429 SKDSTLKVMSYQTKLAQELFGHADEVFGVDWAPDGRSVASGGKD 473
XX
XX RESULT 15
XX ID ABB62260
XX AC ABB62260 standard; protein; 411 AA.
XX ABB62260;
```

```
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 13572.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL06363.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 13572; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 411 AA;
XX
XX Query Match 13.4%; Score 365; DB 4; Length 411;
XX Best Local Similarity 22.7%; Pred. No. 2.5e-25;
XX Matches 110; Conservative 72; Mismatches 155; Indels 148; Gaps 12;
XX
XX QY 6 DEWNLVYRYLOESGFSHSAFTGIESHTS---QSNINGALVPPAALISIIQKGLQYVEA 62
XX Db 11 EELNQAIADYLGNSGYADSLSTFRKADLSTVEKFKGLLEKKWTSVIRLCKKWLEEA 70
XX
XX QY 63 EYSINEDGTLFDRPIESLSLIDAYMPDVVQTRQOAYRDKLAQQAHAATAATNQG 122
XX Db 71 KL-----TEAEK 77
XX
XX QY 123 SAKNGENTANGENGGAHTIANNHTDMEVDGVEIPSKAVVLRGHESEVFICAWNPVSD 182
XX Db 78 EVIEGAPTCKKKTPEGW-----IPRPPK-FSLTGHRSATITVIEHPIFA 121
XX
XX QY 183 LLVSGSGDSTARINWLS-----ENSTSGPTQLVLRHCIREGGQDVPSPKNDVTSLDWNSG 238
XX Db 122 LMVASEDATIRINDFETGETEYERSLKHTDSV-----QDVA-----FDAQK 163
XX
XX QY 239 LIATGSDYDGFARW--TKDGNLASTLGQHKGPIFALKWKKGNFILSAGVDKTTIINDAH 296
XX Db 164 LLASCSADLSIKLWDFQQSIECIKTMGHGHDHNVSVFVDPAGDYVLSRSDTIKQWEVA 223
XX
XX QY 297 TGEAKQFPFPHSPALPDVDMQSNNTFASCTDMCIHVCKLQDPRPIKTFQGHNEVNAIK 356
XX Db 224 TG-----YC-----VKTYTGHEWVRVVR 242
```

```

QY 357 WDPTGNLLASCDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSP-----TGPCTNN 408
Db 243 VHIEGSIATCSNDQIRVWLNTNSKCKVELRDREHTVECIAMAPEAAASAINAEAGADN 302
QY 409 PNAN---LMLASASFDSTVRLWDVDRGICIHITLKHQEPVYVSFAFSDGCRYLASGSFDC 465
Db 303 KKGHHQGPFLASGSRDKTIRIWDVSVGLCLLTLGHDNWNVRGLAFHFGGKYLVSASDDKT 362
QY 466 VHIWN 470
Db 363 IRVWD 367

```

Search completed: August 9, 2004, 16:45:35
Job time : 57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:41:43 ; Search time 13.3333 Seconds
(without alignments)
3708.183 Million cell updates/sec

Title: US-09-987-701-2
Perfect score: 2726
Sequence: 1 MSISSEVNFVRYLQESG.....GDKVGASDSGVCLDLRK 514

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: Piri.*
2: Piri2.*
3: Piri3.*
4: Piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	502.5	18.4	535	2 S48268	probable membrane
2	479	17.6	1526	2 AC2239	WD-40 repeat prote
3	463.5	17.0	1258	2 AI2155	WD-repeat protei
4	460	16.9	1356	2 T18521	beta transducin-li
5	445.5	16.3	1683	2 AF2071	WD-40 repeat prote
6	439.5	16.1	1189	2 AI2433	WD-repeat protei
7	439	16.1	1227	2 AE1810	WD-repeat prote
8	437	16.0	1708	2 AE1866	WD-40 repeat prote
9	414	15.2	1711	2 AD1842	WD-40 repeat prote
10	412.5	15.1	934	2 AG1889	WD-40 repeat prote
11	411.5	15.1	1747	2 AC1842	WD-40 repeat prote
12	390.5	14.3	564	2 T40893	WD repeat protei
13	388.5	14.3	1693	2 S76086	beta transducin-li
14	382.5	14.0	676	2 AB2195	hypothetical prote
15	382	14.0	1551	2 AB2410	WD-repeat protei
16	371.5	13.6	502	2 T41148	trp-asp repeat con
17	368.5	13.5	304	2 AG1837	WD-40 repeat prote
18	368.5	13.5	1189	2 AH1254	WD-repeat protei
19	364.5	13.4	515	2 S19487	hypothetical prote
20	360	13.2	265	2 AF1890	WD-repeat protei
21	350.5	12.9	677	2 AE1861	serine/threonine k
22	348.5	12.8	437	2 S05357	hypothetical prote
23	344.5	12.6	410	2 S48052	platelet-activatin
24	344.5	12.6	559	2 AB2202	hypothetical prote
25	339.5	12.5	786	2 AG2375	WD-40 repeat-prote
26	337	12.4	409	2 S36113	LIS-1 protein - hu
27	335	12.3	333	2 C85034	probable WD-repeat
28	333	12.2	323	2 T02617	hypothetical prote
29	330.5	12.1	473	2 T33805	hypothetical prote

30	329	12.1	798	2 S34023	TATA box-binding p
31	326.5	12.0	787	2 T00798	hypothetical prote
32	324	11.9	342	2 AE2490	WD-repeat protei
33	323.5	11.9	777	2 T41075	hypothetical WD-te
34	321	11.8	376	2 T19266	hypothetical prote
35	320.5	11.8	589	2 AG2400	WD-repeat protei
36	319.5	11.7	606	2 T08180	PF20 protein, micr
37	317.5	11.6	579	2 T22703	hypothetical prote
38	310.5	11.4	704	2 S33263	transcription init
39	309.5	11.4	357	2 AI2099	WD-40 repeat prote
40	309	11.3	1049	2 T42045	beta transducin-li
41	309	11.3	1191	2 S76414	beta transducin-li
42	308	11.3	640	2 S49932	MET10 protein - ye
43	305.5	11.2	317	2 T48032	WD-10 repeat regul
44	304	11.2	605	2 T38932	probable sulfur me
45	300	11.0	701	2 T16607	hypothetical prote

ALIGNMENTS

RESULT 1

S48268
probable membrane protein YBR103w - Yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YBR0832
C;Species: Saccharomyces cerevisiae
C;Date: 01-Aug-1995 #sequence revision 01-Sep-1995 #text_change 19-Apr-2002
C;Accession: S48268; S45971; S44683
R;Manhaupt, G.; Stucka, R.; Ehnl, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A;Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A;Reference number: S48255; MUID:95208357; PMID:7900426
A;Accession: S48268
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-535 <MAN>
A;Cross-references: EMBL:X78993; NID:g476045; PIDN:CAA5606.1; PID:g476059
R;Feldmann, H.; Manhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45927
A;Accession: S45971
A;Molecule type: DNA
A;Residues: 1-535 <FE2>
A;Cross-references: EMBL:Z35972; MIPS:YBR103w
C;Genetics:
A;Gene: SGD:SIF2
A;Cross-references: SGD:S0000307
A;Map position: 2R
C;Superfamily: unassigned WD repeat proteins; WD repeat homology
C;Keywords: transmembrane protein
F;355-388/Domain: WD repeat homology <WD1>
F;397-429/Domain: WD repeat homology <WD2>
F;434-450/Domain: transmembrane #status predicted <TMM>

Query Match 18.4%; Score 502.5; DB 2; Length 535;
Best Local Similarity 24.9%; Pred. No. 3.3e-28;
Matches 142; Conservative 100; Mismatches 232; Indels 97; Gaps 15;

QY	1	MSISSEVNFVRYLQESGFSHSA-----FTFGIKSHISQSNINGALVPPAAL	49
Db	1	MSITSEELNYLIWRYCQEMGHEVSALALQDETRVLEFDEKYKEHI-----PLGTL	50
QY	50	ISLIQKQYVBAEVSINEDG---TLFDCRPISLSLIDAVMPDVVQTRQQAYRDKLAQQ	106
Db	51	VNLVQRGILYSELSMVDKSGDISALNEHLSSEDFNLVQALQID-----KSKFPPI	101
QY	107	QAAAAAASQCSAKNGENTANGAHTIANHHTDMMEVDGVEIPPNKAVVLR	166
Db	102	SSEGRFTLETNSKNKAGEDGASTVERTEQDDTNSISSD--DLGGFKI--LKEIV--	155
QY	167	GHESVFTCAWNPVSD--LLAGSGDSTARINWLSNENSTSGSTQ-----LVLHRCIRGG	219
Db	156	-KLDINVSSTWNPFLDESILAYGEKNSVARLARIVETDQEGKYWKLTIIAELRHPFALSA	214

220 QDVPSNKDVTSLDWNSEGLTATGSDYGFARIWTKDGNLSTLQHGKPIFALKWKKGN 279
215 SSGKTNQVTCVLAHSDGNSIVTGVEGELRLWNTGALLNVLNHFHAPIVSVKWKDGT 274
280 FILSAGVDKTTIWDHATGEAKQFPP-----HSAP---ALDWDQSNNTFAS 324
275 HISMDEVNTILMNVISGTMQHFELKETGSSINAENHSGDGLGVDVWVDDXFEVI 334
325 CSTDMCHVCKLQDRPKITFGHTNEVNAIKWDPNTGNLLASCDMTLKIWSMKQNCV 384
335 PGPKAIFVYITKTPGKLIHGHGPIVSVLEFNDTKLLSASDDGTLRIWHGCGNSQ 394
385 HDLQOHKEIYTIKWSPTGPTNPNANMLASAFDSTVRLWDVDRGICHTLTKEQEP 444
395 NSFVGHSGSIVASV-----GDKVITSMDGVSRLWSLQNTLLALSIVDGPV 444
445 VYSVAFSPDGRYLAGSPDKCHW-----NTQTAL-----VHSYR 481
445 IFAGRISQGGKYAVAFMDGQVNVYDLKLNKSKRSLSYGNRDRGILNPLPIPLVASYQSQ 504
482 GTGGIFECVWNAAGDKVGA--SASDGSVCVL 510
505 DNDYIFDLSWNCAGNKISVAISLQEGSVVAI 535

RESULT 2
AC2239
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2239
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2239
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1526 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075165.1; PID:g17132599; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3466

Query Match 17.6%; Score 479; DB 2; Length 1526;
Best Local Similarity 29.8%; Pred. No. 6.8e-26;
Matches 117; Conservative 76; Mismatches 149; Indels 50; Gaps 10;

QY 136 NGAHTIANHDMEDVDGVEIPFNKAV-VLGEHSEVFCIAWNPVSDLLASGSDSTAR 194
DB 1127 NGV-TLANGSDQIVRLWD--ISSKKCLYTLQHTNWNVAVAFPDGATLASGSDQTVR 1183
QY 195 IWNLSNS-----TSGSTQLVLEHCHREGQDVPSNKDVTSLDWNSEGLTATGSDYGFAR 250
DB 1184 LNDISSKCLYTLQHTSW-----VNSVFNPDGSTLASGSDQTVR 1225
QY 251 IW-TKGNLSTLQHGKPIFALKWKKGNFILSAGVDKTTIWDHATGEAKQFPFHA 309
DB 1226 LWEINSSKCLCTFOGHTSWNSVFNPDGSMASGSDKTVRLWDISSKCLHTFQGH- 1284
QY 310 PALDWDQSNNTF-----ASCSTDMCHVCKLQDRPKITFGHTNEVNAIKWDPNTGN 362
DB 1285 -----NWNVSVAFNPDGSMASGSDQIVRLWEISSKCLHTFQHTSMVSVFSPDGT 1339
QY 363 LLASCSDDMTLKIWSKQNCVHDLQOHKEIYTIKWSPTGPTNPNANMLASAFDS 422
DB 1340 MLASGSDQTVRLWSTSSGECYTLFHTNWNVSVFSPDG-----AILASGSDQ 1390
QY 423 TVRLWDVDRGICHTLTKEQEPVYSVAFSPDGRYLAGSPDKCHWIMNTQTGALVHSYR 482

DB 1391 TVRLWSISSGKCLYTLQGHNNWVGSIFFPDGTLASGDDQTVRLWNISSGECYTLHG 1450
QY 483 -TGGIFECVWNAAGDKVAGASDGSVCVLDLR 513
DB 1451 HINSVRSVAFSSDGLILASGSDDETILKWDVK 1482

RESULT 3
AI2155
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AI2155
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2155
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1258 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074499.1; PID:g17131893; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2800

Query Match 17.0%; Score 463.5; DB 2; Length 1258;
Best Local Similarity 31.8%; Pred. No. 6.7e-25;
Matches 103; Conservative 44; Mismatches 100; Indels 77; Gaps 7;

QY 165 LRGEHSEVFCIAWNPVSDLLASGSDSTARINWLSNSTSGSTQLVLRHCIRREGQDVPS 224
DB 974 LEHTDFTIIGIAFSDPSQTLASASTDSSVRLWNI---STGQCFQILLH----- 1019
QY 225 NKDVTSLDW-----NSEGLTATGSDYGFARIWTKDGNLSTLQHGKPIFALKWKK 277
DB 1020 -----TDWYVAVVFPQGIATGSDCTVKLWNIISTQCCLKTILSEHSDKILGNWSPD 1073
QY 278 GNFIISAGVDKTTIWDHATGEAKQFPFHA PALDWDQSNNTFASCTDMCHVCKL 337
DB 1074 GOLLASASADQSVRLWD-----CCTGECVGI----- 1099
QY 338 QDRPKITFGHTNEVNAIKWDPNTGNLLASCDMTLKIWSMKQNCVHDLQOHKEIYTI 397
DB 1100 -----LRGHSNRVYSAIFSPNGEIIATCSTDQTVKIWDWQGGKCLKTLTGHNTWVFI 1152
QY 398 KWSPTGPTNPNANMLASAFDSTVRLWDVDRGICHTLTKEQEPVYSVAFSPDGRYL 457
DB 1153 AFSPDGK-----ILASAGHDQTVRIWDVNTGKCHHCICHTLHVSVAFPDGEVV 1203
QY 458 ASGSFDCVCHIMNTQTGALVHSYR 481
DB 1204 ASGSQDQTVRIWNVKTEGCLQILR 1227

RESULT 4
T18521
beta transducin-like protein - Podospora anserina
C:Species: Podospora anserina
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18521
R:Saue, S.; Turcq, B.; Begueret, J.
Gene 162, 135-139, 1995
A:Title: A gene responsible for vegetative incompatibility in the fungus Podospora anseri
A:Reference number: Z18944; MUID:96009891; PMID:7557402
A:Accession: T18521
A:Status: preliminary; translated from GB/EMBL/DD5J
A:Molecule type: DNA
A:Residues: 1-1356 <SAU>
A:Cross-references: EMBL:L28125; NID:g607002; PID:g607003; PIDN:AAA85775.1
C:Genetics:

A;Gene: het-e1
A;Introns: 761/3

Query Match
Best Local Similarity 16.9%; Score 460; DB 2; Length 1356;
Matches 124; Conservative 73; Mismatches 178; Indels 70; Gaps 12;

QY 78 IESLSLDVAMPDVVQTRQYQKLAQQQAAAAAASQGGAKNGE-----NT 130
DB 820 ISTISVEAEWNACTQT-----LEHGSSVLSVAFSADQQRVASGSDDKTIKIWD 870
QY 131 ANGENGAHTIANNHTDMEVDGVEIPPNKAVLVRHGESEVFIKAWNPVSDLLASGSD 190
DB 871 ASG--TGTQT-----LEHGSSVLSVAFSADQQRVASGSD 904
QY 191 STAIWNLSENSTGS--TQVLRHCIREGGQDVPSNKDVLSLWNSGTTLLATGSDGPA 249
DB 905 KTIKIW-----DASGTCQTGLEH-----GGR-----VOSVAFSPDQQRVASGSDHTI 949
QY 250 RIW-TKDNGLASTLGOHKGPFIKWKNNKGNFILSAGVDKTTIWDATGEAKQPPFHS 308
DB 950 KIWDAASGTCQTGLEHGGSVLSVAFSPDQQRVASGSDKTIKIWDATASGTCQTGLEHG 1009
QY 309 APALVDWQ--SNTFASCSTDMCIHVCKLQDPRIKTFQHTNEVNAIKWDPNCLLAS 367
DB 1010 GSVMSVAFSPDQQRVASGSDKTIKIWDATASGTCQTGLEHGGVWSVAFSPDQQRVASG 1069
QY 368 SDDMTLKIWSMKQNCVHDILQOHNKIYTIKWSPTGPTNPNANMLASASFDSTVRLW 427
DB 1070 SDDHTIKIWDVSGTCQTGLEHGGDSVWSVAFSPDQ-----RVASGSDIGTIKIWI 1120
QY 428 DVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSYRGTG-I 486
DB 1121 DAASGTCQTGLEHGGVWSVAFSPDQQRVASGSDKTIKIWDAAASGTCQTGLEHGGWV 1180
QY 487 FEVCWNAAGDKVGASADSGVCLVD 511
DB 1181 QSVAFSPDQQRVASGSDKTIKIWD 1205

RESULT 5
AF2071
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AF2071
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1683 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA073823.1; PID:g17131215; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all1214

Query Match
Best Local Similarity 16.3%; Score 445.5; DB 2; Length 1683;
Matches 109; Conservative 69; Mismatches 165; Indels 40; Gaps 6;

QY 124 AKNGENTANGENGGAHTIANNHTDMEVDGVEIPPNKAVLVRHGESEVFIKAWNPVSD 183
DB 1290 SSDKALASASRDNTIKLWNRHGLE-----TFTGHSGGVAVNPLPDSNI 1336
QY 184 LASGSGSTARIWNLSENSTGSQTLVLRHCIREGGQDVPSNKDVLSLWNSGTTLLATG 243
DB 1337 IASASLNTIKLWQRPILSP-----EVLNAGSGVYAVSFLHDSIIATA 1381

QY 244 SYDGFARIW-TKDNGLASTLGOHKGPFIKWKNNKGNFILSAGVDKTTIWDATGEAKQ 302
DB 1382 GADGNIQLWHSQDGSLLKTLPGNKA-IYGISFTPOGDLIASANADKTVKIWRVROGKALK 1440
QY 303 QPFFHSAPALVDWQ--SNTFASCSTDMCIHVCKLQDPRIKTFQHTNEVNAIKWDP 361
DB 1441 TLIGHDNEVKNVNFSPDGTKTILASASRDNTIKLWNSVSGDKTKTLKGHTDEVEFWVSFSDG 1500
QY 362 NLASCSDDMTLKIWSMKQNCVHDILQOHNKIYTIKWSPTGPTNPNANMLASASFD 421
DB 1501 KIILASASADKTIIRLWDSFSGNLIKSLPAHNDLVISVNFNPDGS-----MLASTSAD 1551
QY 422 STVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSYR 481
DB 1552 KTVKLWRSHDGHLLHTFSGHSNVYSSFSFSPDGRYIASASEDKTVKIWDIGHLLTLPQ 1611
QY 482 GTGGIFEVCWNAAGDKVGASASD 504
DB 1612 HQAGVMSAIFSPDGTILISGSLD 1634

RESULT 6
AI2493
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AI2493
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2493
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1189 <KUR>
A;Cross-references: GB:BA000020; PIDN:BA078213.1; PID:g17135667; GSPDB:GN00180
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr7129
A;Genome: plasmid

Query Match
Best Local Similarity 16.1%; Score 439.5; DB 2; Length 1189;
Matches 109; Conservative 65; Mismatches 141; Indels 61; Gaps 9;

QY 167 GHESEVFIKAWNPVSDLLASGSDSTARIWNLSE-----NSTSGSTQLVLRHCIREGGQ-- 220
DB 686 GHDACVMSVVPHPVPGIILATAGEDNTIKLWE-QSGCCLKTLQGHQHWKTIAPNSGGRIL 745
QY 221 -----DVPSNK-----DVTSLDNSEGTLLATGSDYDGFARIW-TKDN 257
DB 746 ASGSPDQNVKLDIHTGKCVMTLQGHGTGVTVSVANPNKDNLLSGSYDQSVKWDRTKR 805
QY 258 LASTLGOHKGPFIKWKNNKGNFILSAGVDKTTIWDATGEAKQFPFHSAPALDV--D 315
DB 806 CLDTLKKNIRLWSVAFHPQGHLPVSGGDDHAAKIWEIWTGQCIKTFQGHSNATYIAH 865
QY 316 WOSNNTFASCSTDMCIHVCKLG-----ODRIKTFQHTNEVNAIKWDPNCLLAS 366
DB 866 WE-HSLLASGHEDQTIKLDNLHSPHKSNNVNTHPRILOGHSHNRVSVFVSFTGQLLAS 924
QY 367 CSDDMTLKIWSMKQNCVHDILQOHNKIYTIKWSPTGPTNPNANMLASASFDSTVRL 426
DB 925 GSADRTIKWSPHGTGCLTLHGHGSWWVAIAFSLD-----DKLLASGSYDHTVKI 975
QY 427 WVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSYRGTG 486
DB 976 WDVSSGQCLQTLQGHGPGSVLAFAVSCDGTLPSSGVEKLVKQWDVETGYCLQT----- 1028
QY 487 FEVCWNAAGDKVGASA 502

Db 1029 ----WEADSNRWAVA 1040

RESULT 7

AE1810
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AE1810
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE1810
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1227 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077553.1; PID:g17135007; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0029

Query Match 16.1%; Score 439; DB 2; Length 1227;

Best Local Similarity 25.4%; Pred. No. 3.7e-23;

Matches 122; Conservative 83; Mismatches 157; Indels 118; Gaps 12;

Db 151 VGDVEI---PNKAV-VLRGHESEVFICAWNPVSDLLASGSGSTARIWNLSE---NS 202

622 MNGEIRLQWTDNKKQIRYKHTAWWAFSPDSRMLASGSAUSTIKLWDVHTGECLKT 681

203 TSGTQLVLRHCIREGGQVPS-NKDVTSLDW-----NSEGT----- 238

682 LSKNTKVVSVAFSPDGRILASQDQTIKLWDIATGNCQQTILGHDDWVMSVTFSPVTD 741

239 ----LLATGSDVGPARIW-TKQGNLSTLGHKGFIPALKNKGNFILSAGVDKTIILW 293

742 DRPLLASSADQIKLMDVATGKCLTKGHTREVSVSPDGTLLASGEDSTVRLW 801

294 DAHTGEAKQPFPSAPALVDWQ-SNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEV 352

802 DVKTCQWQIEFGHSKVVSVRSPDGTQLTASCGEDRSIKLWDIQRGECVNTLWGHSSQV 861

353 NAIKWDPTGNLLASGSDMTLKIVSMKQDNCVHDLQOHKEIYTIKWSP----- 401

862 WAIAPSPDGRITLSCSDQOTARLMDVITGNSLIRGYTRDYVSVAFSPDPSQILASGRDD 921

402 -----TGPGTN----- 407

922 YTIGLNWLKTGECHPLRGHQIRSVAFPHDPCKILASGSADNTIKLWDISDTHSKYIRT 981

408 -----NPNANMLASFPSTVRLWVDVDRGICHTLTGHOEPVSVAFSPDG 454

982 LTGHTNWNWTVVSPDKH-TLASSSEDRITRLWPKD*GDCLQKLKGHSHWVTVAFSPDG 1040

455 RYLASGSPDKVHWTNTOTGALVHSYRGTCG-IFEVCKNAAGDKVGSASDGSVCVLDLR 513

1041 RILASGSADSEIKIWDVASGKCLQTLTDPOGMVSVAFSLDGTLLASGEDQTVKLNLR 1100

RESULT 8

AE1866
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AE1866
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AE1866

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1708 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA072436.1; PID:g17129823; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all0478

Query Match 16.0%; Score 437; DB 2; Length 1708;

Best Local Similarity 30.8%; Pred. No. 8.1e-23;

Matches 119; Conservative 69; Mismatches 154; Indels 44; Gaps 13;

Db 132 NGEENGAHTIANNHTDMWEVDGDIPIPNKAVV---LRGHESEVFICAWNPVSDLLASGS 188

1322 NGETIGSASI-----DATLKWSPQGLLGLTGHNSWVNSVSPDGRIFASGS 1371

189 GDSTARIWNLSENSTSGTQLVLRHCIREGGQVPSNKDVTSLDWNSGTLTATGSDGPF 248

1372 RDKTIVLWDE-----VLLRNPKGDG-----NDWVTSISFSSDGETLAASRDQT 1417

249 ARITWKDGNLSTLGHKGFIPALKNKGNFILSAGVDKTIIWDAHTGEAKQGFPHS 308

1418 VKLSRHGKLLNTFKHTGSIWGVASPNRQMIASAKDQTVKLWH-QDGKILHTLQHQ 1476

309 APALDVDMOSNN-TFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWPTGNMLASC 367

1477 DAVLAVANSSDSQVIASAGKDKIVKWSQG-CQLLHTLQGHTDVAVNVSFSPDGKLLASV 1535

368 SDDMTLKIVSMKQDNCVHDLQOHKEIYTIKWSPGPGTNPNANMLASASDSTVRLW 427

1536 SDDTVKLWS-RDGLLHTLKEHSRRVNGVAVSPDQ-----ILASASIDGTVKLM 1585

428 DVDGRGICHTLTGHOEPVSVAFSPDGRYLASGSPDKVCKVHWTNTOTGALVHSYRG-TGGI 486

1586 NRD-GSLSRNLPDGDGDSFISVSFSPDGKMLAANSDDQ-IRLWN-QKGTLMLVKGDKDEL 1642

487 FEVCWNAAGDKVGSASDGSVCVLDL 512

1643 TSVTFSPSQILAVGGNGKVFILNL 1668

RESULT 9

AD1842

WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AD1842

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AD1842

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1711 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA077808.1; PID:g17135262; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all0284

Query Match 15.2%; Score 414; DB 2; Length 1711;

Best Local Similarity 25.5%; Pred. No. 3.6e-21;

Matches 126; Conservative 74; Mismatches 178; Indels 116; Gaps 15;

Db 65 SINEGDTL----FDGRPIESLIDAVMPDVVQTFQQAAYRKLQAQAAAAAASQ 120

1246 SASDDGTIRLWLDGRPLTIP-----SHTKQ-----VLAVTFSP 1280

121 QGSAXNGENTANGENGAAHTIANNHTDMWEVDGDIPIPNKAVVLRGHESEVFICAWNPV 180

Db 1281 DQQT-----IVSAGADNTVKLSRNTLL-----TTLEGHNEAVMQVIFSPD 1322
QY 181 SLLASGSDGTARLWNLSENSTSGSTQLVLRLHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
Db 1323 GRLIATAGDKTITLWSRDN-----IL-----GTFAGNHVNSLSFSPGNIL 1367
QY 241 ATGSDYGFARITWKGDNASTLGQKGFIPALKWNKNGFILGASGVKDTTIIWDA-----295
Db 1368 ASGSDNTVRLWNTVRLTPKTFYGHKGSVSVRESNDGKITSLSLSTSTWKTWSLQKLL 1427
QY 296 -----HT-----GEAKQOPPHSAPALDVMQSN- 319
Db 1428 QTLASPLDVTISFTPDNPKIVALASPDHTIHLVNRQGLRLSLPGHWHWITSLSFSPNK 1487
QY 320 NTFASCTDMCIHVCKLQDQRPITFOQHTNEVNAIKWDPTGNLLASCSDMTLKIWSMK 379
Db 1488 QILASGSADTKILWSV-NGRLLKTLHGNGWVTDIKFSADGKNIVSASADTKIKWSL- 1545
QY 380 QDNCHVDLQCHNKSIYIKSPGTPGNTNPNANLMLASASPDSTVRLWDVDRGICHTLT 439
Db 1546 DGRILTLQGHASVWSVNLSPDQ-----TLASTSQDSTIKLNLN-CELLYTLR 1595
QY 440 KHOEPVSVAFSPDGRVLASGSPKCVHIWNTOTGALVHSYRG-TGGIPEVCWNAAGDKV 498
Db 1596 GHSVDVNLSPDGTIASASDDGTIKLWNPNGTLLKTFQHRGGRSVSPDQKIL 1655
QY 499 GASASDGSVCVLDL 512
Db 1656 ASGGHDTTVKWNL 1669

RESULT 10
AG1889
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1889
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1889
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1747 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072622.1; PID:g17130010; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0664

Query Match 15.1%; Score 412.5; DB 2; Length 934;
Best Local Similarity 23.7%; Pred. No. 2e-21;
Matches 131; Conservative 113; Mismatches 196; Indels 113; Gaps 21;
QY 8 VNFVYRYLQESGSHAFYGIKSHISQSNINGALVPPAALISIIQKLOYVEAEVSN 67
Db 231 ISSVLTGLIALGAGVAVLQGGKARMSYTK-----AISSAESFLANLEF- 277
QY 68 EDGLTFD---GRPIESLIDA-VMPDVVQTRQAY-----RDKLAAQQAQAAAAA 118
Db 278 -DGLIASIRAGRIKGTGIDANTRTQITETLQCSINFVREKNRLAEHDCMLESVSFSPD 336
QY 119 SQQSSAKNGENTANGEAHTANNHTDMVEVDGVEIPPNKAVVLRGHESEVP-ICAW 177
Db 337 SKFTATASRDKTVK-----IWSJGGKKQL-----VWREXEGEGFNVSAP 376
QY 178 NPVSDLLASGSDGTARLWNLSENSTSGSTQLVLRHCIREGGQ--DVPSNKD-VTSLDWN 234
Db 377 SPDGTLMATGSDWNTAKIWS-----REGKRLHTLDGHEAVLEAVS 418
QY 235 SEGTLTATGSDGFARITWKGDNASTLGQKGFIPALKWNKNGFILGASGVKDTTIIWDA 294

Db 419 PDSQLLATASHDNTVVKLSREGKLLHTLEGHKKNVSIIFSPDQQLIATVGVWNTWKLMN 478
QY 295 AHTGEAKQOPPHSAPALDVMQSN-----SNNTFASCTDMCIHVCKLQDQRPITFOQHT 349
Db 479 L-DGKELRTERFAGH-----DMSVSVSPDQKQIATASGDRTVKLSL-DGKELQTLRGHQ 532
QY 350 NEVNAIKWDPTGNILASCSDMTLKIWSMKQDNCHVDLQCHNKSIYIKWSPTGPG-----405
Db 533 NGVNSVTFSPDGLIATASGDRTVKLSWKSQOB-LETLYCHTDAVNSVAFSPDQTSIATA 591
QY 406 -----TNNPNANLM-----LASASPDSTVRLWDV--DRGICI 435
Db 592 GNDKTAIKLWNLSPNSIIVRGHEDEVDFLVSPNGKYIATASWDKAKLSIVGDKLQEL 651
QY 436 HTLKHOEPVSVAFSPDGRVLASGSPKCVHIWNTOTGALVHSYRG-TGGIPEVCWNA 494
Db 652 RFPNGHQORVVKLSFSPDGIATTSWDKTAKLWNL-DGLQKTLTGKDKTWSVNFSPD 710
QY 495 GDKVGSASDGSV 507
Db 711 QQLIATASEDKTV 723

RESULT 11
AG1842
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1842
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1747 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077807.1; PID:g17135261; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0283

Query Match 15.1%; Score 411.5; DB 2; Length 1747;
Best Local Similarity 27.4%; Pred. No. 5.5e-21;
Matches 114; Conservative 67; Mismatches 156; Indels 79; Gaps 10;
QY 165 LRHSESVFTICANVPVSDLLASGSDGTARIMWL-----SENSTSGSTQ 208
Db 1310 LTGHERITSVKFSPDGKILASASGDKTIKFWNTDGKFLKTLAAHNQVNSINFSSDSKT 1369
QY 209 LV-----LRHCIREGGQV-----PSNK-----226
Db 1370 LVSGADSTMVWKMDGTLLIKTISGRGEQIRDVTFSPDNKVLASASDKTIVRLQNYOK 1429
QY 227 ----DVTSLDWNSEGTLLATGSDYDGFARITWKGDLA-----STLQHKGPFPALKWNKKG 278
Db 1430 SQKSNVNSVFNPDGKTFASAGWDGNTIWRQRE-TLASLSLSTIQKNQNIITVSYPDG 1488
QY 279 NFILSAGVDKTTIWDATGEAKQOPPHSAPALDVMQSNNTFASCTDMCIHVCKLQ 337
Db 1489 KTIATASADNTIKLWDSQTOQLIKTLTGKDR-TTILSPHPDNTIAGSGADTKIWRVN 1548
QY 338 QDRPKTFQHTNEVNAIKWDPTGNLLASCSDMTLKIWSMKQDNCHVDLQCHNKSIYI 397
Db 1549 DGQLRLTLTGHNDEVTSVNFSPDGQFLASGSTDNTVKIW-QTDGLRIKNTIGHGLAISV 1607
QY 398 KWSPTGCTNNPNANLMLASASPDSTVRLWDVDRGICHTLTGHOEPVSVAFSPDGRYL 457
Db 1608 KFSFD-----SHTLASASWNTIKLQVTDGKLIINLNCHIDGVTLSLSPDGEIL 1658

QY 458 ASGSFDCVHIWNTOTGALVHSYRG-TGGEVVCNNAAGDKVGSASDGSVCVLDL 512
 Db 1659 ASGSADNTIKLWNPATLLKTLHGPKINTLAFSPDGKTLGSGEDAGVWVWNL 1714
 RESULT 12
 T40883
 WD repeat protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T40883
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z21954
 A:Accession: T40883
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-564 <WOO>
 A:Cross-references: EMBL:AL031764; FIDN:CAA21113.1; GSPDB:GN00068; SPDB:SPCC1235.09
 A:Experimental source: strain 972h-; cosmid c1235
 C:Genetics:
 A:Gene: SPDB:SPCC1235.09
 A:Map position: 3
 A:introns: 18/1; 273/3; 413/3
 Query Match 14.3%; Score 390.5; DB 2; Length 564;
 Best Local Similarity 24.5%; Pred. No. 3.7e-20;
 Matches 133; Conservative 107; Mismatches 201; Indels 101; Gaps 24;
 QY 3 ISSDEVNLYRYLOESGFHSARTFGIKHISQSNIN---GALVPPAALISIIKGLQY 59
 Db 1 MDTNQVNYIIWRYLKEGYSHTKFAFERETGI--QNLDKQMGSTCOVGALVEILQKGLQY 58
 QY 60 VEAEV-----SINEDG--TLFQGRPIESLSID-----AVMPDVVQT----- 94
 Db 59 VELEKHVYDHNHSENEASKSIDG-----SLVNEPCKLPFYLTVPHCITTLTKADST 113
 QY 95 -----RQAYRDKLAQQOAAAAAASQGSAGKNG-----ENTANGEENGHTIA 142
 Db 114 NGFCEHNNSNDHQLKIQDGGSGSPSPVMPFKDKTEKRDITWADSNVEKDPARPIA 173
 QY 143 NNHTDMVEVDGVEIIPNKAVLRLGHE--SEVF-----ICA-WNPVSD-----LLA 185
 Db 174 VYNSSPV-----TEITIKQVTTGGEDIKSDPFKVPKHPVTCDMPRLQENYHYVE 228
 QY 186 SSGSDSTARLWNL-----ENSTSGTQLVLRHCIRGGQDVPNSKNDVTSLDNSEGTLLA 241
 Db 229 FSIQMTNATLASVSIQEEQNDFRAKTD---YCL-----QSSFDNQDITGVANNSSGFLA 280
 QY 242 TGSYDGFARTWDXGNLSTLGHKGPILFALKWKNKGNFILSAGVDKTTIWDATHTGEAK 301
 Db 281 YAFSFGVIEYDSHGSGQLSFHNKGPVLSLWSGTDITYLAAGSADGTTILFD---QLK 336
 QY 302 Q-QPPEH--SAPALDWDWOSNNTFASCSTDMCIHVCKLGQDRPIKTF-QGHNEVNAIKW 357
 Db 337 QTCYSIDTLASSVLDIEMISDFEFTSDVEGSLRVYKVDGKAPVSTVSHAHNSIVALRY 396
 QY 358 DPTGNLLASCDMDTKIWSMKQD---NCVHDLQHNKEIYTIKWSPTGPGNPNANLM 414
 Db 397 NLRISLLTASSDTTVKLSRSGDAGAFGLH-VFSSFPVNCIDW-----NLREGTPI 448
 QY 415 LASASPDSTVRLWDVDRGICHTLTKHQBPVYVAFSPDGRVLSGSPDKCVHIWNTOTG 474
 Db 449 LAVAS-NSIVSMYNALSLQQLAVFMHTAPVLSLSFHNRYLATGDTSGGVCIIWSCRTA 507
 QY 475 AL 476
 Db 508 XL 509

RESULT 13

S76086

beta transducin-like protein, 190K - Synechocystis sp. (strain PCC 6803)

N;Alternate names: protein sil0163
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 26-May-2000
 C:Accession: S76086
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S76086
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1693 <KAN>
 A:Cross-references: EMBL:D63999; GB:AB001339; NID:G1001396; PIDN:BA010064.1; PID:d1010711
 C:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology
 F:1051-1084/Domain: WD repeat homology <WD01>
 F:1092-1125/Domain: WD repeat homology <WD02>
 F:1133-1166/Domain: WD repeat homology <WD03>
 F:1174-1207/Domain: WD repeat homology <WD04>
 F:1256-1289/Domain: WD repeat homology <WD05>
 F:1297-1330/Domain: WD repeat homology <WD06>
 F:1338-1371/Domain: WD repeat homology <WD07>
 F:1420-1453/Domain: WD repeat homology <WD08>
 F:1461-1494/Domain: WD repeat homology <WD09>
 F:1502-1535/Domain: WD repeat homology <WD10>
 F:1584-1617/Domain: WD repeat homology <WD11>
 F:1625-1658/Domain: WD repeat homology <WD12>
 Query Match 14.3%; Score 388.5; DB 2; Length 1693;
 Best Local Similarity 28.5%; Pred. No. 2.3e-19;
 Matches 107; Conservative 52; Mismatches 138; Indels 79; Gaps 10;
 QY 100 RDKLAQQOAAAAAASQGSAGKNGENTANGE--ENGAHTIANNHTDMVEYDGDVEI 157
 Db 1362 RDKTARLWTTTEGECVAVLADHOGWVREGQFSPDGQWIVTGS---ADKTAQLWNVLG--- 1414
 QY 158 PPNKAVLVGHSEVFFICANNFVSDLLASGSDSTARINWLSNSTSGSTQLVLRHCIRE 217
 Db 1415 ---KKLTVLRGHQDVLNVRFSPDSQYIVTASKQGTAEVWN-----NTGRELAVLRH--- 1463
 QY 218 GGQDVPNSKQVTSLDNKGSEGTLLATGSDGFARITWDXGNLSTLGHKGPILFALKWKNK 277
 Db 1464 -----YEKNIFAEPFADGQFIVTASDDNTAGIWEIVGREVCICRGEHGVVFAQFSAD 1517
 QY 278 GNFLSAGVDKTTIWDATHTGEAKQPPFHSAPALDWDWOSNNTFASCSTDMCIHVCKLG 337
 Db 1518 SRYILTASVDNTARIWDF-----LG 1537
 QY 338 QDRPIKTFQGHTEVNAIKWDPNTGNLLASCSDDMTLKIWSMKQDNCVHDLQHNKEIYTI 397
 Db 1538 --RPLLLTAGHSQSVIYQARFSPGEGNLIATVSADHTARLWD-RSGKTVAVLYGHQGLVGT 1594
 QY 398 KWSPTGTGNTNPNANMLASAFDSTVRLWDVDRGICHTLTKHQBPVYVAFSPDGRYL 457
 Db 1595 DWSPDGQ-----MLVTASNDGTARLWDLIS-GRELLTLEGHGNWVRSAAEFSPDGRWV 1644
 QY 458 ASGSFDCVHIWNTOT 473
 Db 1645 LTSSADGTAKLWPKVT 1660
 RESULT 14
 AH2195
 hypothetical protein alr3119 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AH2195
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2195
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-676 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA874818.1; PID:gl7132214; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3119

```
Query Match      14.0%; Score 382.5; DB 2; Length 676;
Best Local Similarity 23.6%; Pred. No. 1.8e-19;
Matches 123; Conservative 85; Mismatches 195; Indels 119; Gaps 12;

QY 21 FHSATFTGKSHISOSNGALVPPAALISITQKGLQVVEA-----EVS 65
DB 226 FSLGATCFHLLTGTFNSL-----FVEQGSVSWESWQOYWNITSNDREGEYL 273
QY 66 INEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRKLAQQQAAAAA-----AAAAASQOG 122
DB 274 VKVNLKLETDIORRYQSADEVVNDLTKQSLLSRLKTTIPKSAIFRSWSASTSLTAST 333
QY 123 SAKNGENTANGEANGAHTIANNHTDMVEVDG-----DVEIPPKNKAVVLRG 167
DB 334 TKQAKWLNGLRK-QOLLINTMSALLGVGVGHLSQLPLQITKFSIESTQP----YTLKG 389
QY 168 HESEVFICAWNPVSDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSNKD 227
DB 390 HASDVNSVAFSPNGEFLASGSDDKTIKVMNL-----420
QY 228 VTSLDWNSEGTLLATGSYDFARIWTKDGNLSTLQGHKGPFIALKWKKGNFILSAGVD 287
DB 421 -----KNKOKIHTLPCHSGWVAIAFSPDGKTLASTGAD 454
QY 288 KTTIINDAHTGEAKQPPFHSAPALVDWO-SNNTFASCTDMCIHVCKLGODRPIKTFQ 346
DB 455 KTIKLNWLNATGKEIRHLKGHSQGVASVAFSPDGKTLASGSLDKTIKLNPNATGKEIRTLQ 514
QY 347 GHTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGPT 406
DB 515 EHSQGVANVAFSPDGKTLASGSDKTIKLNWLTTSKVIHTLKGHSDLVMSVAFN-----568
QY 407 NNPANMLASASFDSTVRLWDVDRGICHTLTKHOEPVYVAFSP-DGRYLASGSFDCX 465
DB 569 ---SDSQTLAGSKDKTIKLNWLTSTGKTIPTLRGHSKDNVAVVPADSTVLAGSGSDNT 625
QY 466 VHIWNTOTGALVHSY-RGTGCI FEVCWNAAGDKVGASADGS 506
DB 626 IKLWNLTTGEIIRTKRDSGYIVSIVISPDGRNL---ASGGS 664
```

RESULT 15

AB2410
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AB2410
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2410
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1551 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA876533.1; PID:gl7133971; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr4834

```
Query Match      14.0%; Score 382; DB 2; Length 1551;
Best Local Similarity 22.6%; Pred. No. 6.1e-19;
Matches 122; Conservative 94; Mismatches 193; Indels 130; Gaps 18;

QY 48 ALISLIQKGLQVVEAEYSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRKLAQQQ 107
DB 903 ALMSAMRSG-----KALQALVKGDSGLAKYPATSPLLALQITLIDNIQERNQ-----FQGHQ 953
QY 108 AAAAAAASQOGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPKNKAVVLRG 167
DB 954 AWRVSVSFSDRGQVILTASDDCTAR-----LWNLQG-----KQLISLQG 992
QY 168 HESEVFICAWNPVSDLLASGSDSTARIWNLSNSTSGSTQLV-----RHCIR 216
DB 993 HEDTIWISANFSPDGKVIATASDSTRNLWNFSGOOLAKFQGHQGVRSVSFSPDGKHIAT 1052
QY 217 EG-----GQDV-----PSNKDV-TSLD-----WNSGTL-- 240
DB 1053 AGDDHTAELWFSGOQLVQFPFHSGHQTVMCISFSPDGKHIAAADRIVRLWNLKGLLVR 1112
QY 241 -----ATGSYDGPARIWTKDGNLSTLQGHKGPFIALKWKKGNF 280
DB 1113 FPGHQDQVMDVSFSDSQVIATASDSTRNLWNLAGEQITFRGHQGVVSVFSPNGQY 1172
QY 281 ILSAGVDKTTIINDAHTGEAKQPPFHSAPALVDVWQSNNTF-ASCSTDMCIHVCKLGOD 339
DB 1173 IATSSDRTARWNLN-QQLAQSFGHODYVRSVSFSDGKVIATASDSTRVRLWHLNQ 1231
QY 340 RPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIK 399
DB 1232 Q-FSAFQGHQSTVRSVDFSPDGQKVTAAADRTVRLWNLKGEELIQFL-GHRGKWSVSF 1289
QY 400 SPTGPTNNPNANMLASASFDSTVRLWDVDRGICHTLTKHOEPVYVAFSPDGRIYAS 459
DB 1290 SPDGK-----VIATTSDDRTVRLWDI-TGQLLOQFPQGHQGVRSVSFSDGHIAT 1339
QY 460 GSFDKCVHIMNTQTGALV-----HSYRGTCGIFEVWNAAGDKVG 499
DB 1340 ASSDLTTLRLSLDGGELMQFKGHDKWVYVSFSCNGQHIATAADDTARLWNLAGRQVG 1398
```

Search completed: August 9, 2004, 16:48:48
Job time : 15.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:41:13 ; Search time 34.6667 Seconds
(without alignments)
4678.164 Million cell updates/sec

Title: US-09-987-701-2
Perfect score: 2726
Sequence: 1 MSISDEVNFLVRYLQESG.....GRKVGASADGVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL.25.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriaph.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2691	98.7	514	11 Q8BHJ5	Q8bhj5 mus musculus
2	2688	98.6	514	11 Q8CBQ4	Q8cbq4 mus musculus
3	2687	98.6	514	11 Q9E0D4	Q9eqd4 mus musculus
4	2644.5	97.0	519	13 Q7SZM9	C7szm9 xenopus lae
5	2484.5	91.1	527	11 Q8BVC4	Q8byd4 mus musculus
6	2480.5	91.0	527	11 Q8BMM0	Q8bmm0 mus musculus
7	2437	89.4	577	4 Q8EUY2	Q8euy2 homo sapien
8	1972.5	72.4	412	11 Q8COA1	Q8co01 mus musculus
9	1904.5	69.9	700	5 Q9XZK1	Q9xzk1 drosophila
10	1387	50.9	613	10 Q9FN19	Q9fn19 arabidopsis
11	1096	40.2	201	11 Q8VEG3	Q8veg3 mus musculus
12	923.5	33.9	524	5 Q9SRU9	Q9srj9 drosophila
13	477	17.5	1356	3 Q8X1P4	Q8xlp4 podospora a
14	474	17.4	1356	3 Q8X1P5	Q8xlp5 podospora a
15	472	17.3	1356	3 Q8X1P3	Q8xlp3 podospora a
16	449	16.5	1376	3 Q8X1P2	Q8xlp2 podospora a

17	439.5	16.1	1189	16 Q8YL09	Q8yl09 anabaena sp
18	439	16.1	1227	16 Q8Z0R1	Q8z0r1 anabaena sp
19	437	16.0	1708	16 Q8Y2I2	Q8y2i2 anabaena sp
20	414	15.2	1711	16 Q8Z0I9	Q8z0i9 anabaena sp
21	412.5	15.1	934	16 Q8Y2Z3	Q8y2z3 anabaena sp
22	411.5	15.1	1747	16 Q8Z0Z0	Q8z0z0 anabaena sp
23	390.5	14.3	564	3 O74845	O74845 schizosacch
24	382.5	14.0	676	16 Q8YSG6	Q8ysg6 anabaena sp
25	382	14.0	1551	16 Q8YMU3	Q8ymu3 anabaena sp
26	376	13.8	415	4 Q8N136	Q8n136 homo sapien
27	372	13.6	1233	17 Q8TWX4	Q8tmx4 methanosarc
28	371.5	13.6	502	3 O74855	O74855 schizosacch
29	370.5	13.6	481	5 Q8VPR4	Q8vpr4 drosophila
30	370.5	13.6	488	5 Q8T4A2	Q8t4a2 drosophila
31	370	13.6	415	4 Q8N776	Q8n776 homo sapien
32	368.5	13.5	304	11 Q9D4T2	Q9d4t2 mus musculu
33	368.5	13.5	304	16 Q8Z054	Q8z054 anabaena sp
34	368.5	13.5	1189	16 Q8YTD1	Q8ytd1 anabaena sp
35	368	13.5	411	5 Q8G698	Q8g698 drosophila
36	364	13.4	1241	2 Q9XBD8	Q9xbd8 amycolatops
37	362.5	13.3	480	5 Q96995	Q96995 drosophila
38	361.5	13.3	410	13 Q803D2	Q803d2 brachydanio
39	361	13.2	476	13 Q93531	Q93531 xenopus lae
40	361	13.2	476	13 Q7ZXK9	Q7zxk9 xenopus lae
41	360	13.2	265	16 Q8Y2I6	Q8y2i6 anabaena sp
42	359	13.2	352	10 Q80990	Q80990 arabidopsis
43	358.5	13.2	1430	16 Q98HK1	Q98hk1 rhizobium l
44	355.5	13.0	410	13 Q90ZL4	Q90z14 xenopus lae
45	354	13.0	339	11 Q9DCZ7	Q9dcz7 mus musculu

ALIGNMENTS

RESULT 1

Q8BHJ5 ID Q8BHJ5 PRELIMINARY; PRT; 514 AA.
AC Q8BHJ5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE IRA1 protein.
GN IRA1 OR 803049H02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK029595; BAC28241.1; -
DR EMBL; AK033347; BAC28241.1; -
DR PUR; PT0651; PT0651
DR MGD; MGI:2441730; Iral
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PFO0400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00936; Lish; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F8BF14 CRC64;

Query Match 98.7%; Score 2691; DB 11; Length 514;
 Best Local Similarity 98.6%; Pred. No. 6.2e-179;
 Matches 507; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSISSDEVNFIYVRYLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQVY 60
 DB 1 MSISSDEVNFIYVRYLOESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQVY 60
 QY 61 EAEVSIINEDGTLFQGRPIESLSLIDAVMPDVQTRQAYRDKLAQOQAAAAAASQ 120
 DB 61 EAEVSIINEDGTLFQGRPIESLSLIDAVMPDVQTRQAYRDKLAQOQAAAAAATNQ 120
 QY 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVILRGHSESEVFIQWNPV 180
 DB 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVILRGHSESEVFIQWNPV 180
 QY 181 SDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSSEGLL 240
 DB 181 SDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSSEGLL 240
 QY 241 ATGSYDGFARITWKGNIASLTGQHKGPITFALKWKKGNFILSAGVDKTTIINDAHTGEA 300
 DB 241 ATGSYDGFARITWKGNIASLTGQHKGPITFALKWKKGNFILSAGVDKTTIINDAHTGEA 300
 QY 301 KQQPPFHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 DB 301 KQQPPFHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSKQDNCVHDLQHNKEIYTIKWSPTGPTGNPNANMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSKQDNCVHDLQHNKEIYTIKWSPTGPTGNPNANMLASASF 420
 QY 421 DSTVRLWDVDRGICITHTLTKHQBEPVYSAFSPDGRYLASGSPKCVHIWNTQTALVHSY 480
 DB 421 DSTVRLWDVDRGICITHTLTKHQBEPVYSAFSPDGRYLASGSPKCVHIWNTQTALVHSY 480
 QY 481 RGTGGIFEVCNNAAGDKVGSASDGSVCVLDLRK 514
 DB 481 RGTGGIFEVCNNAAGDKVGSASDGSVCVLDLRK 514

RESULT 2
 OSCB64 PRELIMINARY; PRT; 514 AA.
 ID OSCB64
 AC OSCB64
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE IRAL protein.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RX MEDLINE=22354683; Pubmed=12466851;
 RA The FANTOM Consortium,
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK036064; BAC29294.1;
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SMO0667; Lish; 1.
 DR SMART; SMO0320; WD40; 8.

DR PROSITE; PS00896; Lish; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PSS0082; WD_REPEATS_2; 6.
 DR PROSITE; PSS0294; WD_REPEATS_REGION; 1.
 SQ SEQUENCE 514 AA; 55689 MW; 13BED3753A725029 CRC64;

Query Match 98.6%; Score 2688; DB 11; Length 514;
 Best Local Similarity 98.4%; Pred. No. 1e-178;
 Matches 506; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSISSDEVNFIYVRYLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQVY 60
 DB 1 MSISSDEVNFIYVRYLOESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQVY 60
 QY 61 EAEVSIINEDGTLFQGRPIESLSLIDAVMPDVQTRQAYRDKLAQOQAAAAAASQ 120
 DB 61 EAEVSIINEDGTLFQGRPIESLSLIDAVMPDVQTRQAYRDKLAQOQAAAAAATNQ 120
 QY 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVILRGHSESEVFIQWNPV 180
 DB 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVILRGHSESEVFIQWNPV 180
 QY 181 SDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSSEGLL 240
 DB 181 SDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSSEGLL 240
 QY 241 ATGSYDGFARITWKGNIASLTGQHKGPITFALKWKKGNFILSAGVDKTTIINDAHTGEA 300
 DB 241 ATGSYDGFARITWKGNIASLTGQHKGPITFALKWKKGNFILSAGVDKTTIINDAHTGEA 300
 QY 301 KQQPPFHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 DB 301 KQQPPFHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSKQDNCVHDLQHNKEIYTIKWSPTGPTGNPNANMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSKQDNCVHDLQHNKEIYTIKWSPTGPTGNPNANMLASASF 420
 QY 421 DSTVRLWDVDRGICITHTLTKHQBEPVYSAFSPDGRYLASGSPKCVHIWNTQTALVHSY 480
 DB 421 DSTVRLWDVDRGICITHTLTKHQBEPVYSAFSPDGRYLASGSPKCVHIWNTQTALVHSY 480
 QY 481 RGTGGIFEVCNNAAGDKVGSASDGSVCVLDLRK 514
 DB 481 RGTGGIFEVCNNAAGDKVGSASDGSVCVLDLRK 514

RESULT 3
 Q9EQD4 PRELIMINARY; PRT; 514 AA.
 ID Q9EQD4
 AC Q9EQD4
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE IRAL.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RA Zhang X.; Dormay S.; Basch R.;
 RT "Identification of four human cDNAs that are differentially expressed
 RT by early hematopoietic progenitors."
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF268195; AAG44738.1;
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINRPT.

DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; L18H; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; L18H; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 514 AA; 55689 MW; 6A72CE68A40C141F CRC64;

Query Match 98.6%; Score 2687; DB 11; Length 514;
Best Local Similarity 98.4%; Pred. No. 1.2e-178;
Matches 506; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSISDEVNFLVRYLOESGFHSATFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db |||||
QY 1 MSISDEVNFLVRYLOESGFHSATFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db |||||
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOAAAAAASQ 120
Db |||||
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOAAAAAASQ 120
Db |||||
QY 121 QGSAKNGENTANGEENGATIANHNTDMMEVDGVEIPNKAVLGRHSESEVFI 180
Db |||||
QY 121 QGSAKNGENTANGEENGATIANHNTDMMEVDGVEIPNKAVLGRHSESEVFI 180
Db |||||
QY 181 SLLASGSDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSKDVTSLDWSEGITL 240
Db |||||
QY 181 SLLASGSDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSKDVTSLDWSEGITL 240
Db |||||
QY 241 ATGSYDGFARIWTKDGNLASTLGQHKGPFIKWKNGKGNFILLSAGVDKTTIIWDAHTGEA 300
Db |||||
QY 301 KQGFPHSPALVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDP 360
Db |||||
QY 301 KQGFPHSPALVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDP 360
Db |||||
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKEIYTIKWSPTGPTGNPNANMLASAF 420
Db |||||
QY 421 DSTVRLWDVDRGICHTLTTHKQEPVYSVAFSPDGRYLASGSPDKCVHIWNTQTGA 480
Db |||||
QY 481 RGTGGIFECVWNAAGDKVGASDGSVCVLDLRK 514
Db |||||
481 RGTGGIFECVWNAAGDKVGASDGSVCVLDLRK 514

RESULT 4
Q7SZM9 PRELIMINARY; PRT; 519 AA.
ID Q7SZM9 AC Q7SZM9
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DE Nuclear receptor co-repressor complex subunit TBLR1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xeropodinae; Xenopus.
OC NCBI_TaxID=8395;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12794076;
RA Tomita A., Buchholz D.R., Obata K., Shi Y.B.;
"Fusion Protein of Retinoic Acid Receptor (alpha) with Promyelocytic
Leukemia Protein or Promyelocytic Leukemia Zinc Finger Protein
RT Recruits N-Cor-TBLR1 Corepressor Complex to Repress Transcription in
RT Vivo.";
RL J. Biol. Chem. 278:30788-30795 (2003).

DR EMBL; AY225088; AAP20646.1; --
KW Receptor.
SQ SEQUENCE 519 AA; 56043 MW; 5B998BDC8C892296 CRC64;

Query Match 97.0%; Score 2644.5; DB 13; Length 519;
Best Local Similarity 96.3%; Pred. No. 1.1e-175;
Matches 500; Conservative 4; Mismatches 10; Indels 5; Gaps 1;

QY 1 MSISDEVNFLVRYLOESGFHSATFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db |||||
QY 1 MSISDEVNFLVRYLOESGFHSATFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db |||||
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOAAAAAASQ 118
Db |||||
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOAAAAAASQ 120
Db |||||
QY 119 --SQQSAGKNGENTANGEENGATIANHNTDMMEVDGVEIPNKAVLGRHSESEVFI 175
Db |||||
QY 121 TPNQQQPAKNGENTANGEENGATIANHNTDMMEVDGVEIPNKAVLGRHSESEVFI 180
Db |||||
QY 176 AWPVSDLLASGSDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSKDVTSLDWNS 235
Db |||||
QY 181 AWPVSDLLASGSDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSKDVTSLDWNS 240
Db |||||
QY 236 EGTLLATGSDYDGFARIWTKDGNLASTLGQHKGPFIKWKNGKGNFILLSAGVDKTTIIWDA 295
Db |||||
QY 241 EGTLLATGSDYDGFARIWTKDGNLASTLGQHKGPFIKWKNGKGNFILLSAGVDKTTIIWDA 300
Db |||||
QY 296 HTGEAKQGFPHSPALVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAI 355
Db |||||
QY 301 HTGEAKQGFPHSPALVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAI 360
Db |||||
QY 356 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKEIYTIKWSPTGPTGNPNANML 415
Db |||||
QY 361 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKEIYTIKWSPTGPTGNPNANML 420
Db |||||
QY 416 ASASFDSTVRLWDVDRGICHTLTTHKQEPVYSVAFSPDGRYLASGSPDKCVHIWNTQTGA 475
Db |||||
QY 421 ASASFDSTVRLWDVDRGICHTLTTHKQEPVYSVAFSPDGRYLASGSPDKCVHIWNTQTGA 480
Db |||||
QY 476 LVHSYRTGIGIFECVWNAAGDKVGASDGSVCVLDLRK 514
Db |||||
481 LVHSYRTGIGIFECVWNAAGDKVGASDGSVCVLDLRK 519

RESULT 5
Q8BYQ4 PRELIMINARY; PRT; 527 AA.
ID Q8BYQ4 AC Q8BYQ4
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE Transducin (Tbalx protein).
GN TBLX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hypothalamus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

```

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton B., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schermer A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RC SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; AK038674; BAC30092.1; -
DR ENBL; BC043105; AH43105.1; -
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS50896; Lish; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
SQ SEQUENCE 527 AA; 56802 MW; 146435A9C51DFDA6 CRC64;

Query Match 91.1%; Score 2484.5; DB 11; Length 527;
Best Local Similarity 87.9%; Pred. No. 1.5e-164;
Matches 463; Conservative 28; Mismatches 21; Indels 15; Gaps 2;

QY 1 MSISDDEVNLFVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSITSDEVNLFVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQQAAYDKLAQQOAAAAA----- 115
DB 61 EAEISINEDGTVFDGRPIESLSLIDAVMPDVVQTRQQAAYDKLAQQOAAAAA----- 120
QY 116 -----AAASQGSAGKNGENTANGENGATIANHTDMMEVDGVEIPNKAVALR 166
DB 121 ATSTAATPAAAAQONPKNGEATVNGEENGAAHAI -NNHSPKMEIDGVEIPPSKATVLR 179
QY 167 GHSEVEFICAWNPVSDLLASGSDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNK 226
DB 180 GHSEVEFICAWNPVSDLLASGSDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNK 239
QY 227 DVTSLDWNSDGLTATGSDYGFARITWKGMLASTLGQHGKPIFALKWNKKNFILLSAGV 286
DB 240 DVTSLDWNSDGLTATGSDYGFARITWKGMLASTLGQHGKPIFALKWNKKNFILLSAGV 299
QY 287 DKTTIIIDAHTGEAKQPPHSPALDWDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQ 346
DB 300 DKTTIIIDAHTGEAKQPPHSPALDWDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQ 359
QY 347 GHTNEVNAIKWDPTGNILLASCSDDMTLKIWSMKQDNCHDLQOHKEIYTIKWSPTGPGT 406
DB 360 GHTNEVNAIKWDPTGNILLASCSDDMTLKIWSMKQDNCHDLQOHKEIYTIKWSPTGPGT 419
QY 407 NNPNANLKLASASPDSTVRLMDVDRGICHTLTKEPVPVSVAFSPDGRYLASGSPDKCV 466

```

```

QY 347 GHTNEVNAIKWDPGTGNLLASCDDMTLKIWSMKQDNCHDLQHNKEIYTIKWSPTGRT 406
Db 360 GHTNEVNAIKWDPGSMGLASCDDMTLKIWSMKQACVHDLQHSKEIYTIKWSPTGRT 419
QY 407 NPNANMLASASFDSTVRLWVDRGICHTLTKHQEPVYVAFSPDGRYLAGSFGDKCV 466
Db 420 SNPNINMLASASFDSTVRLWVDRGICHTLTKHQEPVYVAFSPDGRYLAGSFGDKCV 479
QY 467 HWNTQSGLVHSYRGTTGIFEVCNNAAGDKVGASDGSVCVLDLRK 514
Db 480 HIWNTQSGLVHSYRGTTGIFEVCNNAAGDKVGASDGSVCVLDLRK 527

RESULT 7
Q86UY2 PRELIMINARY; PRT; 577 AA.
AC Q86UY2;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE TBLIX protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.P.,
RA Brownstein M.J., Ustin B., Ioshizuka S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC052304; AAHS2304.1; -.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
SQ SEQUENCE 577 AA; 62495 MW; D830A37781E2A15C CRC64;

```

```

Query Match      89.4%; Score 2437; DB 4; Length 577;
Best Local Similarity 86.0%; Pred. No. 3.4e-161;
Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

```

```

QY 1 MSISSDEVFLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISITQKGLQYV 60
Db 52 MSITSDEVFLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISITQKGLQYV 111
QY 61 EAEYSINEDGLPDRPIESLSLIDAVPDPVOTRQOAVRDKLAQQOAAAAAATAA 117
Db 112 EAEYSINEDGLPDRPIESLSLIDAVPDPVOTRQOAFREKLAQQOAAAAAATAA 171
QY 118 -----ASQGSKANGENTANGENGAHTIANNHTDMMEVDGVEIPPNKAVYLRG 167
Db 172 ATAATTTTSGVSHONPSKNEATVNGENRAHSV-NNHAKPMEIDGVEIPSSKATVLRG 230
QY 168 HSEVEFICANNPYSDLLASGSDSTARINWLSNSISGSGTOLVLRHCIRGGQDVPNSKD 227
Db 231 HSEVEFICANNPYSDLLASGSDSTARINWLSNSISGSGTOLVLRHCIRGGQDVPNSKD 290
QY 228 VTSLDNMSGTLLATGSDYGFARITWKDGNLSTLGHKGPIFALKWNKGNFTLSAGVD 287
Db 291 VTSLDNMSGTLLATGSDYGFARITWKDGNLSTLGHKGPIFALKWNKGNFTLSAGVD 350
QY 288 KTIINDAHTGEAKQPPHSAFALVDVQSNNTFASCSTDMCIHVCKLQODRPIKTFQG 347
Db 351 KTIINDAHTGEAKQPPHSAFALVDVQSNNTFASCSTDMCIHVCKLQODRPIKTFQG 410
QY 348 HTNEVNAIKWDPGTGNLLASCDDMTLKIWSMKQDNCHDLQHNKEIYTIKWSPTGRT 407
Db 411 HTNEVNAIKWDPGTGNLLASCDDMTLKIWSMKQDNCHDLQHNKEIYTIKWSPTGRT 470
QY 408 NPNANMLASASFDSTVRLWVDRGICHTLTKHQEPVYVAFSPDGRYLAGSFGDKCV 467
Db 471 NPNANMLASASFDSTVRLWVDRGICHTLTKHQEPVYVAFSPDGRYLAGSFGDKCV 530
QY 468 HWNTQSGLVHSYRGTTGIFEVCNNAAGDKVGASDGSVCVLDLRK 514
Db 531 HWNTQSGLVHSYRGTTGIFEVCNNAAGDKVGASDGSVCVLDLRK 577

RESULT 8
Q8COAL PRELIMINARY; PRT; 412 AA.
AC Q8COAL;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Transducin (Fragment).
GN TBLIX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK031937; BAC27612.1; -.
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
FT NON TER 1
SQ SEQUENCE 412 AA; 44356 MW; EB78910E6D9E5237 CRC64;

```

Query Match	72.4%; Score 1972.5; DB 11; Length 412;
Best Local Similarity	87.6%; Pred. No. 4.4e-129;
Matches 360;	Conservative 22; Mismatches 28; Indels 1; Gaps 1;
QY	104 AQQQAAAAAASQSSAKNGENTANGENGARTIANDHMDVEVDGVEIPPKAV 163
DB	3 AATATSTATTAAAAQQPPKNGEATVNGEENGAAH-NNHKKPMEIDGVEIPPKAT 61
QY	164 VLRGHSEVFICAMPVPSDLLASGGSDSTARLWNLSENSTGSLVLRHCIREGGQDVP 223
DB	62 VLRGHSEVFICAMPVPSDLLASGGSDSTARLWNLSENSTGSLVLRHCIREGGHVP 121
QY	224 SNKDVTSLDWNBEGTLATGSDYGFARIMTKDGNLSTLQGHKGFIFALKNNKGNFILS 283
DB	122 SNKDVTSLDWNBEGTLATGSDYGFARIMTKDGNLSTLQGHKGFIFALKNNKGNFILS 181
QY	284 AGVDKTTIWDHTGAKQFPFHSAPALDWDQSNTPFASCTDMCHVCKLGQDRPK 343
DB	182 AGVDKTTIWDHTGAKQFPFHSAPALDWDQSNTPFASCTDMCHVCKLGQDRPK 241
QY	344 TFOGHTNEVNAIKWDPTEGNLLASCSDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTG 403
DB	242 TFOGHTNEVNAIKWDPTEGNLLASCSDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTG 301
QY	404 PGTNNPNAIMLASAFDSFTVLWDVDRGICHTLTKQEPVYSVAFSPDGRVYLASGSGD 463
DB	302 PATSNPNSIMLASAFDSFTVLWDVDRGICHTLTKQEPVYSVAFSPDGRVYLASGSGD 361
QY	464 KCVHIWNTOTGALVHSYRGTTGGIFECVWNAAGDKVGASDGSVCVLDLRK 514
DB	362 KCVHIWNTOTGALVHSYRGTTGGIFECVWNAAGDKVGASDGSVCVLDLRK 412

RESULT 9

Q9XZK1	PRELIMINARY; PRT; 700 AA.
ID	Q9XZK1
AC	Q9XZK1
DT	01-NOV-1999 (T-EMBLrel. 12, Created)
DT	01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE	EBI protein.
GN	EBI OR C04063.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RC	SEQUENCE FROM N.A.
RC	STRAIN=Berkeley.
RC	MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Ananatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Aril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Darbin K., Doup L.E., Dowes M., Dugan-Rocha S., Dunkov B.C., Durn P.,
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Heiman T.J.J., Hernandez J.R., Houck J.,
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA	Jalali M., Kaulush K.A., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
RA	Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA	Reinart K., Renington K., Saunders R.D.C., Scheeler F., Shen H.,
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA	Spiel E., Stadling A.C., Stapleton M., Strong R., Sun E.,
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT	"The genome sequence of Drosophila melanogaster.";
RL	Science 287:2185-2195(2000).
RN	SEQUENCE FROM N.A.
RP	MEDLINE=9234084; PubMed=10215623;
RP	Dong X., Tsuda L., Zavitz K.H., Lin M., Li S., Carthew R.W.,
RA	Zipursky S.L.;
RA	"ebi regulates epidermal growth factor receptor signaling pathways in
RT	Drosophila.";
RL	Genes Dev. 13:954-965(1999).
RN	SEQUENCE FROM N.A.
RP	STRAIN=Berkeley.
RC	Celniker S.E., George R.A., Galle R.F., Hoskins R.A., Svirskas R.R.,
RA	Harris N.L., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA	Chavez C., Chew M., Doyle C.M., Farfan D.E., Flanagan J.,
RA	Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim S.H., Lee B.,
RA	Lomotab M.A., Mak J., Mazda P., Moshrefi A.R., Moshrefi M., Nixon K.,
RA	Pacleb J.M., Park S., Pfeiffer B., Punch E., Snir E., Twomey B.,
RA	Wan K.H., Zhang R., Zieran L., Rubin G.M.;
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AE003589; AAF51501.1; -
DR	EMBL; AF146345; AAD35017.1; -
DR	EMBL; AC005762; -; NOT ANNOTATED_CDS.
DR	FLYBase; FBgn0023444; ebi.
DR	GO; GO:0000074; P:regulation of cell cycle; IMP.
DR	InterPro; IPR006594; Lish.
DR	InterPro; IPR001005; Myb_DNA_binding.
DR	InterPro; IPR001680; WD40.
DR	Pfam; PF00400; WD40; 8.
DR	PRINTS; PR00320; GPROTEINERPT.
DR	ProDom; PD000018; WD40; 3.
DR	SMART; SMC0667; Lish; 1.
DR	SMART; SMC0320; WD40; 8.
DR	PROSITE; PS00896; LISH; 1.
DR	PROSITE; PS00037; MYB_1; 1.
DR	PROSITE; PS00678; WD_REPEATS_1; 3.
DR	PROSITE; PS00082; WD_REPEATS_2; 6.
DR	PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW	Repeat; WD repeat.
SQ	SEQUENCE 700 AA; 72387 MW; 28C6D8D07BB79PB7 CRC64;
Query Match	69.9%; Score 1904.5; DB 5; Length 700;
Best Local Similarity	54.8%; Pred. No. 4.9e-124;
Matches 385;	Conservative 51; Mismatches 75; Indels 191; Gaps 8;
QY	1 MSISDSVNFVLYRLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISITQKGLQYV 60
DB	1 MSFSDSVNFVLYRLOESGFLHSAYVFGIESHSQSNINGALVPPAALTILQKGLYT 60
QY	61 EAVYSINEDGTLFGRIEISLIDAVMPDV-----VQTRQ----- 96
DB	61 EVESVSGEDGEV--ARPIEGLSLIDAVMPVKPLKTVKTEPGKPGAVDSAPAGNQNN 118
QY	97 -----QAYRDKLA-----QQQAAAAAASQSSAKNGENT 130
DB	119 NAKPEIKIEPTGTVAGSAGGNKIASGTGTSTPTDQASAEVDSGSGNANNAGTYAGNNG 178
QY	131 ANGBE----- 135

```
Db 179 AGNQASTGGSNSTSTPACGDLAAPGASQKQSNSEAGSSSGNAGNANATSTDDAASS 238
QY 136 ---NG-----AHTIANNHTEM-----148
Db 239 TSTNGNSSTSSVQPTSGLTGAGTIVTSNPDAASGGASTATSKAPSGAVTIRGQAQ 298
QY 149 -----MEVDGVEIPPNKAVVLRGHESEV 172
Db 299 GNNVQSGSSNAQSSAFSGTSSSTSGGAGTAAALVPMIDENIEIPESKARVLRGHESEV 358
QY 173 FICAMPVSDLLASGSGDSTARWNLSENSTSGSTOLVLRHCIRGGQDVPNSKDVTSID 232
Db 359 FICAMPVSDLLASGSGDSTARWNLSENSTSGSTOLVLRHCIRGGQDVPNSKDVTSID 417
QY 233 WNSEGTLATGSDGFGARITWTKGNLSTLQGHKGFIFALKWKKGNFILSAGVDKTTII 292
Db 418 WNCDSGLLATGSDGVARIWTKTDLGLASTLQGHKGFIFALKWKKGNFILSAGVDKTTII 477
QY 293 WDAHTGEAKQPPFHSAPALDWDQSNNTFASCTDMCIHVCKLQGDRIPIKTFQGHTEV 352
Db 478 WDASTGCTQCFAFHSAPALDWDQSNNTFASCTDMCIHVCKLQGDRIPIKTFQGHTEV 537
QY 353 NAIKMDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQGHKGFIFALKWKKGNFILSAGVDKTTII 412
Db 538 NAIKMDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQGHKGFIFALKWKKGNFILSAGVDKTTII 597
QY 413 LMLASASPDSTVRLWDVDRGICHLITLKHQEPVSVASPDGRYLASGDFKCVHIWNTQ 472
Db 598 LMLASASPDSTVRLWDVDRGICHLITLKHQEPVSVASPDGRYLASGDFKCVHIWNTQ 657
QY 473 TGAHVHSYGTGIGFVPCVNAAGDKVGASASGSCVCLDLRK 514
Db 658 TGAHVHSYGTGIGFVPCVNAAGDKVGASASGSCVCLDLRK 699

RESULT 10
Q9FN19 ID Q9FN19 PRELIMINARY; PRT; 613 AA.
AC Q9FN19;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genomic DNA, chromosome 5, TAC clone:K8K14 (AT5G67320/K8K14_4).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned P1 clones.";
RL DNA Res. 4:401-414(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

```
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hayashizaki Y.,
RA Hsuan V.W., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M.,
RA Southwick A., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yamada K., Yu G., Yuan S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007645; BAB09017.1; -
DR EMBL; AY057698; AAL15328.1; -
DR EMBL; AY143932; AAN28971.1; -
DR InterPro; IPR006594; LISH.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRODOM; PD00018; WD40; 2.
DR PRINTS; PR00320; GPROTEINRPT.
DR SMART; SM00667; LISH; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00882; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ
SEQUENCE 613 AA; 69772 MW; 04F40772311C0E76 CRC64;

Query Match 50.9%; Score 1387; DB 10; Length 613;
Best Local Similarity 44.2%; Pred.No. 4.le-88;
Matches 272; Conservative 91; Mismatches 144; Indels 108; Gaps 7;

QY 2 SISDEYNELVRYLQESGFSHSAFTGKISHLSQSNINGALVPPALISIIKQGLQYVE 61
Db 3 SLTSVEINFLVRYLQESGFTHAFTLGYEAGINKSNIDGNMPPGALIKFVQKGLQYME 62
QY 62 -----AEVSINEDGTLFDGRTIESLSLIDAVMPDVVTRQQAAYRDKLAQCCAAAAA 114
Db 63 MEANLNSSEVDIEDPSFF--QPLDLISKVDKELQDLREKKRDMERKERSKNDK 120
QY 115 AAAASQOQSAGKNGENTANGE-----GDVE-----ENGAHTIAN 143
Db 121 GVEREHEGDRNRAKEKDRHEKQKEREKEREKEREKEREKEREKEREKEREKERE 180
QY 144 NHTDMMEVD-----GDVE-----156
Db 181 REKDLKLEKEREIEREREREKEREKEREKEREKEREKEREKEREKEREKEREKERE 240
QY 157 -----IPNKAVVLRGHESEVFTCANPVSDLLASGSGDSTARINWLSNSTS 204
Db 241 DIVMTPTTSQTSHPNSDVRILEGTSVCACAWSPSASLLASGSGDATARIWISPE 300
QY 205 -----GSTOLVLRHCIRGGQDVPNSKDVTSLDWNSGTLTATGSDGFGARITW 258
Db 301 AVHTGRNINAILKHA---KGKSNESKDVTTLDWNGEGTLTATGSCDQGARITW 357
QY 259 ASTLQGHQGFIPALKWKKGNFILSAGVDKTTIIWDAHTGEAKQPPFHSAPALD 318
Db 358 ISTLSKHGKGFIFALKWKKGNFILSAGVDKTTIIWDAHTGEAKQPPFHSAPALD 417
QY 319 NNTFASCSDDMTLHVCKLQGDRIPIKTFQGHTEVNAIKMDPTGNLLASCSDD 378
Db 418 NVGFATSDTSMIYLCKIGETRPKFTFGQGEVNCVCKWDPTGSLASCSDSDSTAK 477
QY 379 KQCNVCHDLQGHKGFIFALKWKKGNFILSAGVDKTTIIWDAHTGEAKQPPFHS 438
Db 478 KQSTFVCHDLREHTKEIYTIWSPGTGPTGNPNKQLTLASASFSFSTVKLWDAEL 537
QY 439 TKHQEPVSVASPDGRYLASGDFKCVHIWNTQTGALVHSYGTGIGFVPCVNA 498
Db 538 NGHREPVSILAFSPNGEYIASGLSDKSIHWSIKEGKIVKTYTGNGGIFVPCW 597
QY 499 GASASGSCVCLDLR 513
```


Db 598 AACFADNSVCVLDLR 612

RESULT 11

Q8VEG3 PRELIMINARY; PRT; 201 AA.

AC Q8VEG3; (1)

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Similar to IRAL protein (Fragment).

OS Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR ENBL; BC018512; AAH18512.1; -

DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40; 4.

DR ProDom; PD000018; WD40; 2.

DR SMART; SM00320; WD40; 4.

DR PROSITE; PS00678; WD_REPEATS_1; 2.

DR PROSITE; PS00682; WD_REPEATS_2; 3.

DR PROSITE; PS00294; WD_REPEATS_REGION; 1.

KW Repeat; WD repeat.

FT NON TER

SQ SEQUENCE 201 AA; 22038 MW; 6B945F137B491818 CRC64;

Query Match 40.2%; Score 1096; DB 11; Length 201;

Best Local Similarity 99.0%; Pred. No. 1.5e-68;

Matches 199; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 314 VDMQSNNTFASCTDNCIHVKLGQDRPIKTFQGHTEVNAIKWDPKGNLLASCSDDMTL 373

Db 1 VDMQSNNTFASCTDNCIHVKLGQDRPIKTFQGHTEVNAIKWDPKGNLLASCSDDMTL 60

QY 374 KWSMKQDNCVHDLQHNKEIYTIKWSPTGPGNNPNNALMLASASFDSTVRLWDVDRGI 433

Db 61 KWSMKQDNCVHDLQARNKEIYTIKWSPTGPGNNPNNALMLASASFDSTVRLWDVDRGI 120

QY 434 CIHTLTKHQPVPYVAFSPDGRVLAGSPDKVHIWNTQTGALVHSYRGTFGIFVCWNA 493

Db 121 CIHTLTKHQPVPYVAFSPDGRVLAGSPDKVHIWNTQTGALVHSYRGTFGIFVCWNA 180

QY 494 AGDKVGASADSGVCVLDLRK 514

Db 181 AGDKVGASADSGVCVLDLRK 201

RESULT 12

Q95RJ9 PRELIMINARY; PRT; 524 AA.

AC Q95RJ9; (1)

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE L24373P.

GN EBI OR CG4063.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=Berkeley;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.B., Rubin G.M., Celnikier S.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR ENBL; AY061326; AAL28874.1; -

DR FlyBase; FBgn0023444; ebi.

DR GO; GO:0000074; P:regulation of cell cycle; IMP.

DR InterPro; IPR006594; Lish.

DR InterPro; IPR001005; Myb_DNA_binding.

DR Pfam; PF00400; WD40; 3.

DR PRINTS; PR00320; GPROTEINBRPT.

DR ProDom; PD000018; WD40; 1.

DR SMART; SM00667; Lish; 1.

DR PROSITE; PS00896; LISH; 1.

DR PROSITE; PS00037; MYB_1; 1.

DR PROSITE; PS00678; WD_REPEATS_1; 2.

DR PROSITE; PS00682; WD_REPEATS_2; 3.

DR PROSITE; PS00294; WD_REPEATS_REGION; 1.

KW Repeat; WD repeat.

SQ SEQUENCE 524 AA; 52529 MW; 6C3D86110BA18D65 CRC64;

Query Match 33.9%; Score 923.5; DB 5; Length 524;

Best Local Similarity 40.5%; Pred. No. 5.8e-56;

Matches 218; Conservative 39; Mismatches 64; Indels 217; Gaps 12;

QY 1 MSISDEVNLFVRYLQESGFHSFAFTFGIKSHISQSNINGALVPPAALISIIKGLQYV 60

Db 1 MSFSDEVNLFVRYLQESGFHSFAFTFGIKSHISQSNINGALVPPAALISIIKGLQYV 60

QY 61 EAEVINEGTLFDGRPIESLSLIDAMPDV-----VQTRQ----- 96

Db 61 EVENSVDGEV-ARPIEGLSLIDAMPVEKPLPIKTEPKGPGAVDSAPAGNQN 118

QY 97 -----QAYRDKLA----- 130

Db 119 NAKPEIKIEPOTGVAGSAGGNKIAGSTTGTTDQSASEVDSGNANAGTYAGNG 178

QY 131 ANGE----- 135

Db 179 AGNQASTGGSNSTTPAGGLAAPGASQKQNSNEAGSSSGNAGNANATSTDDAAS 238

QY 136 ---NG-----AHTIANNHTM----- 148

Db 239 TSTNGNSTSSVEQPTSGLTPTAGTIVSTNPDAASGGASTATGSKAPSGAVTIRVGAQ 298

QY 149 -----MEVDGDVEIPPNKAVLVRGHESEV 172

Db 299 GNNVQSGSSNAQSSAPSGTISSTSGGAGTTPAALVPMDIDENIPEISKARVLRGHESEV 358

QY 173 FICAWNPSVLLASGSDSTARIWNLSENSTSGTQVLVLRHCIREGGQDYPNSKNDVTSLD 232

Db 359 FICAWNPSRDLASGSDSTARIWMDSDANTN-SNQLVLRHCIRKQGGAEVPSNNDVTSLD 417

QY 233 WNSEGTLLATGSYDGFARIWTKDGNLASTLQGHKGPFIKLNKKNKGFILSAGVDKTTII 292

Db 418 WNCDSLLATGSYDGYARIWTKDGRLASTLQGHKGPFIKLNKKNKGVNLSAGVDKTTII 477

QY 293 WDAHTGKAKQFPFHSAPALDVMQSNNTFASC-----STDNCIHW-CK 335

Db 478 WDASTG-----PMHPAICL-----SQCSILGCGGLADKPGCLLQYGSADTCTCPVPGCK 524

RESULT 13

Q8X1P4

ID Q8X1P4

AC Q8X1P4; PRELIMINARY; PRT; 1356 AA.

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Beta transducin-like protein HET-E2C*4.

GN HET-E.

OS Podospora anserina.


```
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Begueret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of het-e and het-d genes";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323583; AAL37299.1; -
DR InterPro; IPR007111; NACHT_NTPase.
DR Pfam; PF05729; NACHT; 1.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PS08337; NACHT; 1.
DR PROSITE; PS0678; WD_REPEATS_1; 10.
DR PROSITE; PS0082; WD_REPEATS_2; 10.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 1356 AA; 149702 MW; 33148AP4A7B82826 CRC64;

Query Match 17.5%; Score 477; DB 3; Length 1356;
Best Local Similarity 26.2%; Pred. No. 2.5e-24;
Matches 133; Conservative 86; Mismatches 215; Indels 74; Gaps 13;

QY 49 LISIQKGLQYVEAEVSNEDGTLFGRIEISLIDAMPVQTRQAYRDLAQQA 108
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
769 LIAIVRGYRFALSYRMIEKAPLQAYTSALVFAPDTSMIKKIFKKEEPTISTISVEA 828
QY 109 AAAAAAASQGS-----AKNGENTANGENGA-----HTIANHTDMMEVDG--D 154
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
829 ENNACTQTLEHGSSVLSVAFSPDQGVASGDDTKIKIWDASGTCTQTLLEGHGSVWS 888
QY 155 VEIPP-----NKAV-----VLRGHESEVFICAWNPVSDLLASGSDST 192
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
889 VAFSPDQGVASGDDTKIKIWDASGTCTQTLLEGHGSVLSVAFSPDQGVASGDKT 948
QY 193 ARIWNLSENSTGSLVLRHCIREGQDVPSNKNVTSLDWNSGTLTATGSDGFARIV 252
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
949 IKIW---DTASGTCTQTLLEGH-----GG-----SVWSVAFSPDQGVASGDKTIKI 994
QY 253 -TKDGNLASTLGQHKGFIPALKWKNKGNFILSAGVDKTIWDATGEAKQFPFHSAPA 311
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
1053 ----WQSVAFSPDQGVASGSDNHTIKIWDASGTCTQTLLEGHGSVWSVAFSPDQGV 1108
QY 365 ASCSDDMTLKWSMKQDNCVHDLQHNKEIYTKMSPTGPTGNPNANMLASASPDSTV 424
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
1109 ASGSDHTIKIWDASGTCTQTLLEGHGSVWSVAFSPDQ-----RVASGSDGTI 1159
QY 425 RLWVDVRCIHTLTKHQEPVSVAFSPDQGVASGDFKCVHWNVTQGVVHVSRTG 484
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
1160 KIWDASGTCTQTLLEGHGSVWSVAFSPDQGVASGSDTKIKIWDASGTCTQTLLEGH 1219
QY 485 G-IFFECVNAAGDKVGASASDGSVCVLD 511
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
1220 GWHSVAFSPDQGVASGSDNTIKIWD 1247

RESULT 14
Q8X1P5 PRELIMINARY; PRT; 1356 AA.
AC Q8X1P5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Beta transducin-like protein HET-E2C.
GN HET-E.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Begueret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of het-e and het-d genes";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323583; AAL37299.1; -
DR InterPro; IPR007111; NACHT_NTPase.
DR Pfam; PF05729; NACHT; 1.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PS08337; NACHT; 1.
DR PROSITE; PS0678; WD_REPEATS_1; 10.
DR PROSITE; PS0082; WD_REPEATS_2; 10.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 1356 AA; 149720 MW; 9A80777304B361D0 CRC64;

Query Match 17.4%; Score 474; DB 3; Length 1356;
Best Local Similarity 26.3%; Pred. No. 4e-24;
Matches 132; Conservative 86; Mismatches 222; Indels 62; Gaps 12;

QY 49 LISIQKGLQYVEAEVSNEDGTLFGRIEISLIDAMPVQTRQAYRDLAQQA 108
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
769 LIAIVRGYRFALSYRMIEKAPLQAYTSALVFAPDTSMIKKIFKKEEPTISTISVEA 828
QY 109 AAAAAAASQGS-----AKNGENTANGENGA-----HTIANHTDMMEVDG--D 154
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
829 ENNACTQTLEHGSSVLSVAFSPDQGVASGDDTKIKIWDASGTCTQTLLEGHGSVWS 888
QY 155 VEIPP-----NKAV-----VLRGHESEVFICAWNPVSDLLASGSDST 192
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
889 VAFSPDQGVASGDDTKIKIWDASGTCTQTLLEGHGSVLSVAFSPDQGVASGDKT 948
QY 193 ARIWNLSENSTGSLVLRHCIREGQDVPSNKNVTSLDWNSGTLTATGSDGFARIV 252
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
949 IKIW---DTASGTCTQTLLEGH-----GG-----SVWSVAFSPDQGVASGDKTIKI 994
QY 253 -TKDGNLASTLGQHKGFIPALKWKNKGNFILSAGVDKTIWDATGEAKQFPFHSAPA 311
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
995 DTASGTCTQTLLEGHGSVWSVAFSPDQGVASGDDTKIKIWDASGTCTQTLLEGHGW 1054
QY 312 LDVDWQ--SNNTFASCSDDMCIHVCKLQDRPIKTFQGRITNEVNAIKWDPTGNLLASGDD 370
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
1055 WSAVAFSPDQGVASGSDGTIKIWDASGTCTQTLLEGHGWVQSVAFSPDQGVASGDD 1114
QY 371 MTLKWSMKQDNCVHDLQHNKEIYTKMSPTGPTGNPNANMLASASDSTVRLWVD 430
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
1115 HTIKIWDASGTCTQTLLEGHGSVWSVAFSPDQ-----RVASGSDGTIKIWDAA 1165
QY 431 RGICHTLTKHQEPVSVAFSPDQGVASGDFKCVHWNVTQGVVHVSRTG--IFEV 489
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
1166 SGTCQTQTLLEGHGSVWSVAFSPDQGVASGSDTKIKIWDASGTCTQTLLEGHGWHSV 1225
QY 490 CWNAAGDKVGASASDGSVCVLD 511
DB : : : : : : : : : : : : : : : : : : : : : : : : : :
1226 AFSPDQGVASGSDNTIKIWD 1247

RESULT 15
Q8X1P3 PRELIMINARY; PRT; 1356 AA.
AC Q8X1P3;
```

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta transducin-like protein HET-E2C*40.
GN HET-E.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Begueret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of het-e and het-d genes."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF323584; AAL37300.1; -
DR InterPro: IPR007111; NACHT_NTPase.
DR InterPro: IPR001680; WD40.
DR Pfam: PF05729; NACHT; 1.
DR Pfam: PF00400; WD40; 10.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD000018; WD40; 10.
DR SMART: SM00320; WD40; 10.
DR PROSITE: PSS0837; NACHT; 1.
DR PROSITE: PSS0678; WD_REPEATS_1; 10.
DR PROSITE: PSS0082; WD_REPEATS_2; 10.
DR PROSITE: PSS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 1356 AA; 149694 MW; 572A0B034DCB0D21 CRC64;

Query Match 17.3%; Score 472; DB 3; Length 1356;
Best Local Similarity 31.7%; Pred. No. 5.5e-24;
Matches 111; Conservative 59; Mismatches 154; Indels 26; Gaps 7;

QY 165 LRGHESVFICAMNPVSDLLASGSDSTARIMNLSENSTSGTQVLVLRHCIREGGQDVPS 224
Db 921 LEHGSSVLSVAFSPDQQRVASGSDTKIKW---DTASGTGTLEGH---GG----- 968

QY 225 NKDVTSIDMNSGTLATGSDYGFARIW-TKQGNLASTLGQHKGPFPALKWKKGNFILS 283
Db 969 --SWSVAFSPDQQRVASGSDTKIKWDTASGTCTQTLLEGHGNVSVAFSPDQQRVAS 1026

QY 284 AGVDKTTIWDATGEAKQFPFHSAPALDWDQO-SNNTFASCSSTMCIHVCKLGQDRPI 342
Db 1027 GSDKTKIKWDTASGTCTQTLLEGHGWVSVAFSPDQQRVASGSDTKIKWDAASGTCT 1086

QY 343 KTFQGHTEVNAIKWDPITGNLLASCSDDMTLKIWSMKQDNCVHDLQCHNKEIYTIKWSPT 402
Db 1087 QTLLEGHDWQSVAFSPDQQRVASGSDTKIKWDAASGTCTQTLLEGHGDVSVAFSPD 1146

QY 403 GPGTNNNNALMLASAFSDTVRLWDVDRGICHTLTKEQEPVYSVAFSPDGRYLASGSF 462
Db 1147 GQ-----RVASGSDTKIKWDAASGTCTQTLLEGHGWVSVAFSPDQQRVASGI 1197

QY 463 DKCVHIWNTQTGALVHSYRTGG-IFEVCNAAAGDKVGASADGVCVLD 511
Db 1198 DGIKIKWDAASGTCTQTLLEGHGWVSVAFSPDQQRVASGSDNTIKIWD 1247

Search completed: August 9, 2004, 16:47:57
Job time : 37.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:39:47 ; Search time 48 Seconds
(without alignments)
3025.613 Million cell updates/sec

Title: US-09-987-701-4
Perfect score: 2718
Sequence: 1 MSISDEVNFLYRYLQESG.....LPHLVWPLVALIELLVLK 514

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04: *
1: Geneseq1980s: *
2: Geneseq1990s: *
3: Geneseq2000s: *
4: Geneseq2001s: *
5: Geneseq2002s: *
6: Geneseq2003as: *
7: Geneseq2003bs: *
8: Geneseq2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2488	91.5	514	4 AAB95225	Human pro
2	2485	91.4	514	6 ABO07190	Human p53
3	2229	82.0	577	7 ADD14051	Human arc
4	2132	78.4	542	5 ABP41760	Human ova
5	2047	75.3	395	5 ABP51424	Human MDD
6	1721.5	63.3	584	4 ABG21351	Novel hum
7	1715.5	63.1	700	4 ABG60376	Drosophil
8	1137	41.8	208	4 AAO04385	Human pol
9	819	30.1	167	4 ABG21350	Novel hum
10	469	17.3	535	6 ABR52980	Protein s
11	374.5	13.8	414	6 ADA13321	Human int
12	370.5	13.6	411	4 ABG62260	Drosophil
13	367.5	13.5	485	4 AAB68284	Amino aci
14	366.5	13.5	485	4 AAB68516	Human GTP
15	366.5	13.5	485	4 AAB92844	Human pro
16	366.5	13.5	485	5 ABB97306	Novel hum
17	359	13.2	484	4 AAB68282	Amino aci
18	356	13.1	352	3 AAG14893	Arabidops
19	356	13.1	352	3 AAG48119	Arabidops
20	356	13.1	352	7 ADB95040	A. thalia
21	353	13.0	349	3 AAG48120	Arabidops
22	353	13.0	349	3 AAG14894	Arabidops
23	351.5	12.9	409	7 ADB54235	Human pro
24	350	12.9	339	2 AAW55957	Human cia
25	350	12.9	339	2 AAW58888	Human lip

26	350	12.9	339	7 ADE59693	Human Pro
27	350	12.9	339	7 ADE59705	Human Pro
28	350	12.9	339	7 ADE59697	Human Pro
29	350	12.9	339	7 ADE59709	Human Pro
30	350	12.9	339	7 ADE83378	Human Pro
31	350	12.9	339	7 ADE59701	Human Pro
32	350	12.9	339	7 ADE59713	Human Pro
33	350	12.9	339	7 ADE57861	Human Pro
34	349.5	12.9	409	2 AAR70002	OPDE 45 k
35	349.5	12.9	409	7 ADE58488	Rat Prote
36	349.5	12.9	409	7 ADE83492	Rat Prote
37	349.5	12.9	409	7 ADE54233	Rat Prote
38	349.5	12.9	410	2 AAR70005	OPDE 45 k
39	342	12.6	481	4 ABB59486	Drosophil
40	341.5	12.6	479	3 AAY79678	Drosophil
41	339	12.5	704	2 AAW06085	Human TAT
42	339	12.5	704	2 AAW25018	TATA-bind
43	338	12.4	409	2 AAR85868	WD-40 dom
44	338	12.4	450	6 ABB99407	Amino aci
45	338	12.4	451	6 ADA21145	Human sec

ALIGNMENTS

RESULT 1
AAB95225
ID AAB95225 standard; protein; 514 AA.
XX AC AAB95225;
XX DT 26-JUN-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:17352.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-00116126.
XX PR 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300253.
XX PR 11-JAN-2000; 2000JP-00118776.
XX PR 02-MAY-2000; 2000JP-00183767.
XX PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 17352; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 XX Sequence 514 AA;

Query Match 91.5%; Score 2488; DB 4; Length 514;
 Best Local Similarity 99.4%; Pred. No. 2.le-224;
 Matches 469; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSISDEVNVLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
 Db 1 MSISDEVNVLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 QY 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICANPV 180
 Db 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICANPV 180
 QY 181 SLLASGSDSTARINWLSNSTSGTQLVLRHCIREGGQDVPSNKOVTSLDWNSEGLL 240
 Db 181 SLLASGSDSTARINWLSNSTSGTQLVLRHCIREGGQDVPSNKOVTSLDWNSEGLL 240
 QY 241 ATGSDGFARWTKDGNLSTLQHKGPFPALKWKKGNFLLSAGVDKTTIWDATGTA 300
 Db 241 ATGSDGFARWTKDGNLSTLQHKGPFPALKWKKGNFLLSAGVDKTTIWDATGTA 300
 QY 301 KQFPFHSAPALVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 Db 301 KQFPFHSAPALVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTGNFNANMLASAF 420
 Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTGNFNANMLASAF 420
 QY 421 DSTVRLMDVDRGICHTLTKHQPVSVAFPSPDGRYLASGSPKCVHIWNTQ 472
 Db 421 DSTVRLMDVDRGICHTLTKHQPVSVAFPSPDGRYLASGSPKCVHIWNTQ 472

RESULT 2
 ABO07190
 ID ABO07190 standard; protein; 514 AA.

XX ABO07190;
 XX
 XX
 XX 13-AUG-2003 (first entry)

XX Human p53 modifying protein, SEQ ID 150.

XX Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
 KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
 KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
 KW apoptotic disorder; cell proliferation disorder.

XX Homo sapiens.

XX W0200299122-A1.

XX

PD 12-DEC-2002.

XX 03-JUN-2002; 2002WO-US017382.

XX 05-JUN-2001; 2001US-0296076P.

PR 10-OCT-2001; 2001US-0328605P.

PR 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.

XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

XX WPI; 2003-156859/15.

XX N-PSDB; ACD13365.

XX Identifying modulators of the p53 pathway for use in treating apoptotic
 PT or cell proliferation disorders, comprises screening for agents that
 PT modulate activity of a human ortholog of genes that modify the p53
 PT pathway in *Drosophila*.

XX Example 2; Page 458-459; 679pp; English.

XX The invention relates to identifying (M1) a candidate p53 pathway
 CC modulating agent, by contacting an assay system comprising a purified HM
 CC polypeptide (human orthologue of genes that modify the p53 pathway in
 CC *Drosophila*) or nucleic acid with a test agent under conditions, where but
 CC for the presence of the test agent, the system provides a reference
 CC activity, and detecting a test agent-biased activity of the assay system.
 CC Also included are modulating (M2) a p53 pathway of a cell (comprising
 CC contacting a cell defective in p53 function with a candidate modulator
 CC that specifically binds to an HM polypeptide comprising an HM amino acid
 CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
 CC in a mammalian cell (comprising contacting the cell with an agent that
 CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
 CC a disease in a patient (comprising: (a) obtaining a biological sample
 CC from the patient; (b) contacting the sample with a probe for HM
 CC expression; (c) comparing the results with a control; and (d) determining
 CC whether the comparison indicates a likelihood disease). (M1) is useful
 CC for identifying modulators of the p53 pathway. A probe for HM expression
 CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
 CC in a patient, where the cancer has greater than 25 % expression level.
 CC Modulators identified by (M1) are useful in a variety of diagnostic and
 CC therapeutic applications, where disease or disorder prognosis is related
 CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
 CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
 CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
 CC the p53 function of the cell, so that the cell undergoes normal
 CC proliferation or progression through the cell cycle. (M2) and (M3) are
 CC also useful for treating defects in the p53 pathway such as angiogenic,
 CC apoptotic or cell proliferation disorders. The present sequence
 CC represents a human p53 pathway modifying protein

XX Sequence 514 AA;

Query Match 91.4%; Score 2485; DB 6; Length 514;

Best Local Similarity 99.4%; Pred. No. 4e-224;
 Matches 469; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSISDEVNVLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
 Db 1 MSISDEVNVLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 QY 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICANPV 180
 Db 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICANPV 180
 QY 181 SLLASGSDSTARINWLSNSTSGTQLVLRHCIREGGQDVPSNKOVTSLDWNSEGLL 240
 Db 181 SLLASGSDSTARINWLSNSTSGTQLVLRHCIREGGQDVPSNKOVTSLDWNSEGLL 240

QY 241 ATGSYDGFARITWKDGNLSTLQHGKGPFPALKNKNGFILSAGVDKTTIWDHTGEA 300
 DB 241 ATGSYDGFARITWKDGNLSTLQHGKGPFPALKNKNGFILSAGVDKTTIWDHTGEA 300
 QY 301 KQOFFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 DB 301 KQOFFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNQVHDLQHNKEIYTIKWSPTGPGTNNPNNMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNQVHDLQHNKEIYTIKWSPTGPGTNNPNNMLASASF 420
 QY 421 DSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRVLAGSPDKCVHIWNTQ 472
 DB 421 DSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRVLAGSPDKCVHIWNTQ 472

RESULT 3

ID ADD14051

XX AC ADD14051 standard; protein; 577 AA.

XX AC ADD14051;

DT 01-JAN-2004 (first entry)

DE Human src biomarker polypeptide SEQ ID NO:240.

XX predictor set; protein tyrosine kinase activity modulator;

XX protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;

XX gene therapy; drug sensitivity; genetic profile; cancer; human.

XX Homo sapiens.

XX WO2003062395-A2.

PD 31-JUL-2003.

XX 17-JAN-2003; 2003WO-US0001981.

XX 18-JAN-2002; 2002US-0350061P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Huang F, Fairchild CR, Lee FY, Shaw P;

XX WPI; 2003-636735/60.

DR N-PSDS; ADUL4646.

XX New polynucleotides and polypeptides for predicting the activity of
 PT compounds that interact with protein tyrosine kinases and/or protein
 PT tyrosine kinase pathways.

XX Claim 10; SEQ ID NO 240; 139pp; English.

XX The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and
 CC correlating the expression of the markers to the compound's ability to
 CC modulate the activity of the cells; (2) a plurality of cell lines for
 CC identifying polynucleotides and polypeptides whose expression levels
 CC correlate with compound sensitivity or resistance of cells associated
 CC with a disease state; and (3) identifying polynucleotides and
 CC polypeptides that predict compound sensitivity or resistance of cells
 CC associated with a disease state, comprising subjecting the plurality of
 CC cell lines to one or more compounds, analysing the expression pattern of
 CC a microarray of polynucleotides or polypeptides, and selecting
 CC polynucleotides or polypeptides that predict the sensitivity or
 CC resistance of cells associated with a disease state by using the

CC expression pattern of the microarray. The polynucleotides and
 CC polypeptides have cytostatic activities, and can be used in gene therapy.
 CC The polynucleotides and polypeptides are useful in predicting the
 CC activity of compounds that interact with protein tyrosine kinases and/or
 CC protein tyrosine kinase pathways. These may be used in determining drug
 CC sensitivity in patients to allow the development of individualized
 CC genetic profiles which aid in treating diseases and disorders (e.g.
 CC cancer) based on patient response at a molecular level. The present
 CC sequence is used in the exemplification of the present invention.

XX Sequence 577 AA;

SQ Query Match 82.0%; Score 2229; DB 7; Length 577;

Best Local Similarity 85.4%; Pred. No. 5.5e-200;

Matches 414; Conservative 32; Mismatches 25; Indels 14; Gaps 2;

QY 1 MSISDEVNFLVRYLOBSGFSAFTFGIKSHISQSNINGALVPPAALISIIKGLQYV 60
 DB 52 MSITSDEVNFLVRYLOBSGFSAFTFGIESHSQSNINGTLVPPAALISIIKGLQYV 111
 QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVNPDVVTQQAAYRDKLAQQQAAAAA--- 117
 DB 112 EAEISINEDGTVDGRPIESLSLIDAVNPDVVTQQAAYRDKLAQQQASAAAAAATAA 171
 QY 118 -----ASQOGSAKNGENTANGENGHTIANNHTMMVEVDGVEIPNKAIVLRG 167
 DB 172 ATAATTTSAGVSHONPSKNREATVNGEENRAHSV-NNAKPEIDGVEIIPSKATVLRG 230
 QY 168 HESEVFICAMNPVSDLLASGSDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKD 227
 DB 231 HESEVFICAMNPVSDLLASGSDSTARINWLSNSTSGTQVLVLRHCIREGGHVDVPSNKD 290
 QY 228 VTSLDWNSGTLTATGSDYDGFARITWKDGNLSTLQHGKGPFPALKNKNGFILSAGVD 287
 DB 291 VTSLDWNTNGTLTATGSDYDGFARITWEDGNLSTLQHGKGPFPALKNKNGNYILSAGVD 350
 QY 288 KTTIWDHTGEAKQOFFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQ 347
 DB 351 KTTIWDHTGEAKQOFFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQ 410
 QY 348 HTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNQVHDLQHNKEIYTIKWSPTG 407
 DB 411 HTNEVNAIKWDPTGMLLASCDDMTLKIWSMKQDNQVHDLQHNKEIYTIKWSPTG 470
 QY 408 NPANMLASASFDSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRVLAGSPDKCVH 467
 DB 471 NPNSNIMLASASFDSTVRLWDIERGVCTHTLTKHQPVPVSVAFSPDGKYLASGSD 530
 QY 468 IWNTQ 472
 DB 531 IWNTQ 535

RESULT 4

ABP41760

ID ABP41760 standard; protein; 542 AA.

XX AC ABP41760;

DT 22-AUG-2002 (first entry)

DE Human ovarian antigen HNOKM38, SEQ ID NO:2892.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;

XX ovarian cancer; breast cancer; tumour; reproductive system disorder;

XX infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;

XX PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;

XX inflammatory condition; immune disorder; blood disorder;

XX cardiovascular disorder; respiratory disorder; neurological disorder;

XX gastrointestinal disorder; urinary system disorder; drug screening;

XX gene therapy; chromosome mapping; forensic analysis;

XX antibody preparation; cytostatic; immunomodulatory; neuroprotective;

XX antiinflammatory; gynaecological; reproductive.

XX OS Homo sapiens.
 XX PN WO200200677-A1.
 XX PD 03-JAN-2002.
 XX PF 07-JUN-2001; 2001WO-US018569.
 XX PR 07-JUN-2000; 2000US-0209467P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Birse CE, Rosen CA;
 XX DR WPI; 2002-147878/19.
 XX DR N-PSDB; ABQ54837.
 XX PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 PT cancer), immune disorders, cardiovascular disorders and neurological
 PT diseases.
 XX PS Claim 11; SEQ ID NO 2892; 2922pp; English.
 XX CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 542 AA;
 Query Match 78.4%; Score 2132; DB 5; Length 542;
 Best Local Similarity 84.8%; Pred. No. 6.5e-191;
 Matches 395; Conservative 31; Mismatches 26; Indels 14; Gaps 2;
 QY 20 GFHSATFTGKHSISNSNGALVPPAALISIIQKGLQVVEAEVINEEDGLTFDCGRPIE 79
 DB 36 GFHSATFTXIESHISNSNGTLVPPAALISIIQKGLQVVEAEISINEEDGVFDGRPIE 95
 QY 80 SLSLIDAVMDVQTRQAYRDKLAQQAQAAAAA-----ASQQGSAKN 126
 DB 96 SLSLIDAVMDVQTRQQAQREKLAQQAQAAAAAATAATAATTSAGVSHQNPKN 155
 QY 127 GENTANGENCAHTIANHTDMEVDGVEIPNKAIVLREHSEVETICANNPVSLLAS 186
 DB 156 REATVNGENRAHSV-NNHAKPMEIDGVEIPSKATVLRGHESEVETICANNPVSLLAS 214

QY 187 GSGDSTARINWLNSENSTSGTQLVLRHCIREGGDVPNKDVTSLDWNSGTTLATGSYD 246
 DB 215 GSGDSTARINWLNSENSTSGTQLVLRHCIREGGDVPNKDVTSLDWNTNGTLLATGSYD 274
 QY 247 GFARITWTKDGNLSTLQGHKGPPIFALKWKKGNFIISAGVDKTTIIWDAHTGSAKQOFPF 306
 DB 275 GFARITWEDGNLSTLQGHKGPPIFALKWKKGNVILSAGVDKTTIIWDAHTGSAKQOFPF 334
 QY 307 HSAPALVDWQSNNTFASCTDMCIHYCKLGQBPDKTFOGHTNEVNAIKWDPTGNLLAS 366
 DB 335 HSAPALVDWQSNNTFASCTDMCIHYCKLGQBPDKTFOGHTNEVNAIKWDPTGNLLAS 394
 QY 367 CSDDMTLKINSMKQDNCVHDLQCHNKIYTIKWSPTGPTNNPNANLMLASASFDSTVRL 426
 DB 395 CSDDMTLKINSMKQEVCIHDLQAHNKIYTIKWSPTGPTNNPNANLMLASASFDSTVRL 454
 QY 427 WDVDRGICITLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 472
 DB 455 WDIERGVCVTHLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 500
 RESULT 5
 ID ABP51424 standard; protein; 395 AA.
 AC ABP51424;
 DT 03-SEP-2002 (first entry)
 DE Human MDDT SEQ ID NO 446.
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatotropic; antiinflammatory; antipsoriatric; cytostatic; anti-HIV;
 KW antiallergic; antianaemic; antiaesthetic; antithrombotic; antitumor;
 KW neuroprotective; antirheumatic; antiarthritic.
 OS Homo sapiens.
 XX WO200240715-A2.
 XX 23-MAY-2002.
 XX 06-SEP-2001; 2001WO-US027628.
 XX 05-SEP-2000; 2000US-0229747P.
 XX 05-SEP-2000; 2000US-0229748P.
 XX 05-SEP-2000; 2000US-0229749P.
 XX 05-SEP-2000; 2000US-0229750P.
 XX 05-SEP-2000; 2000US-0229751P.
 XX 06-SEP-2000; 2000US-0230583P.
 XX 06-SEP-2000; 2000US-0230584P.
 XX 06-SEP-2000; 2000US-0230515P.
 XX 06-SEP-2000; 2000US-0230517P.
 XX 06-SEP-2000; 2000US-0230518P.
 XX 06-SEP-2000; 2000US-0230519P.
 XX 06-SEP-2000; 2000US-0230595P.
 XX 06-SEP-2000; 2000US-0230597P.
 XX 06-SEP-2000; 2000US-0230598P.
 XX 06-SEP-2000; 2000US-0230599P.
 XX 06-SEP-2000; 2000US-0230610P.
 XX 06-SEP-2000; 2000US-0230685P.
 XX 06-SEP-2000; 2000US-0230988P.
 XX 06-SEP-2000; 2000US-0230989P.
 XX 07-SEP-2000; 2000US-0230951P.
 XX 07-SEP-2000; 2000US-0231163P.
 XX 07-SEP-2000; 2000US-0231167P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX XX

PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;
 PI Hillman JL, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PB;
 PI Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B;
 PI Roseberry AM, Gerstin EH, Petalta CH, David NH, Panzer SR, Flores V;
 PI Daffo A, Marwaha K, Chen AU, Chang SC, Au AP, Inman RR;
 XX WPI; 2002-527544/56.
 DR N-PSDB; AB072641.
 XX
 XX Novel human disease detection and treatment polypeptide, useful in
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.
 PT AIDS.
 XX
 PS Claim 14; Page 578; 618pp; English.
 XX
 XX The invention relates to an isolated human disease detection and
 CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a
 CC sequence selected from 234 sequences (ABP51231-ABP51484) given in the
 CC specification, a naturally occurring polypeptide comprising a sequence
 CC having at least 90% identity to (I) or a biologically active or
 CC immunogenic fragment of (I). (I) is useful for screening a compound for
 CC effectiveness as an agonist or antagonist, for screening a compound that
 CC specifically binds (I) or modulates the activity of (I), and for
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 CC screening a compound for effectiveness in altering expression of a target
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
 CC detecting MDDT in a sample or for assessing toxicity of a test compound,
 CC in a diagnostic test for a condition or a disease associated with the
 CC expression of MDDT in a biological sample, for detecting (I) in a sample,
 CC and for purifying (I) from a sample. A composition comprising (I), an
 CC agonist or antagonist is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional MDDT. (I)
 CC or (II) are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of MDDT, where the disorders are
 CC selected from a cell proliferative disorder such as arteriosclerosis,
 CC cirrhosis, hepatitis, psoriasis, and cancer and an
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised
 CC animals or transgenic animals to model human diseases, in somatic or
 CC germline gene therapy, to generate a transcript image of a tissue or cell
 CC type, for detecting differences in the chromosomal location due to
 CC translocation or inversion among normal, carrier or affected individuals
 CC and as hybridisation probes for mapping naturally occurring genomic
 CC sequences
 XX
 SQ Sequence 395 AA;

Query Match 75.3%; Score 2047; DB 5; Length 395;
 Best Local Similarity 99.5%; Pred. No. 3.8e-183;
 Matches 389; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSISDEVNPLVRYLQESGFHSATFGIKSHISQSNINGALVPPAALISIIQGLQV 60
 Db 5 MSISDEVNPLVRYLQESGFHSATFGIESHISQSNINGALVPPAALISIIQGLQV 64
 QY 61 EAEVSNEDGTLFDPGPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 Db 65 EAEVSNEDGTLFDPGPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 124
 QY 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVLRGHESEVFICAWNPV 180
 Db 125 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVLRGHESEVFICAWNPV 184
 QY 181 SLLAGSGDSTARINWLNSENSTSGTLVLRHCRREGQDVPSKNDVTSLDWNSEGLL 240
 Db 185 SLLAGSGDSTARINWLNSENSTSGTLVLRHCRREGQDVPSKNDVTSLDWNSEGLL 244
 QY 241 ATGSYDGFARIWTKDGNLSTLQGHKGPFLKWNKGNFISAGVDKTTIWDHTGEA 300
 Db 245 ATGSYDGFARIWTKDGNLSTLQGHKGPFLKWNKGNFISAGVDKTTIWDHTGEA 304

QY 301 KQOPFHSAPALDWDWQSNNTFASCSDMCIHVCKLGQDRIKTFQGHTNEVNAIKWDPT 360
 Db 305 KQOPFHSAPALDWDWQSNNTFASCSDMCIHVCKLGQDRIKTFQGHTNEVNAIKWDPT 364
 QY 361 GNLLASCSDMTLKIWSMKQDNCVHDLQOHN 391
 Db 365 GNLLASCSDMTLKIWSMKQDNCVHDLQAHN 395

RESULT 6

ABG21351

ID ABG21351 standard; protein; 584 AA.

XX AC ABG21351;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #21342.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS85538.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 20; SEQ ID NO 51710; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: the sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 584 AA;

Query Match

63.3%; Score 1721.5; DB 4; Length 584;

Best Local Similarity 63.5%; Pred. No. 2.7e-152;
Matches 338; Conservative 30; Mismatches 35; Indels 129; Gaps 4;

QY 16 LQSGFHSFAFTGKSHISQSNINGALVPPAALISIIQKGLQYVEARVINEEDGTLFG 75
DB 62 LEVKGFSFAFTGIESHISQSNINGTLVPPSALISILQKGLQYVEARISINKDGTVPDS 121
QY 76 RPTESLIDAVMPDVQFQYRQKLAQQQAAAAA-----AAASQOGSA 124
DB 122 RPTESLIVAVIPDVVQFQYRQKLAQQQAAAAA-----AAASQOGSA 181
QY 125 KNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFCICAWNPVSDLL 184
DB 182 KREATVNGEAGAEI--NNHSPMEIDGDVEIPPNKATVLRGHESEVFCICAWNPVSDLL 240
QY 185 ASGSGDSTARINWLSNSTSGTQVLVLRHCIRREGGQDVPSNKDVTSLDWSGTLATGS 244
DB 241 ASGSGDSTARINWLSNSTSGTQVLVLRHCIRREGGQDVPSNKDVTSLDWSGTLATGS 300
QY 245 YDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGFIILSAGVDKTTIWDATGKQOF 304
DB 301 YDGFARITW----- 310
QY 305 PPHAPALDVWDGNNNTFASCSTDMCHVCKLQODRPKTFQHT----- 349
DB 311 --NAPALDVWDGNNNTFASCSTDMCHVCKLQODRPKTFQHT----- 367
QY 350 -----NEVNAIKWDPT 360
DB 368 KLTMTGKELRPKTPCSDGALLPPAGRRPHLLTGDFIFFKRLSALFQNEVNAIKWDS 427
QY 361 GNLLASCDSDMTLKIMSKODNCVHDLQOHKEIYTIKWSPTGPGTNNFNANMLASAF 420
DB 428 GMLLASCDSDMTLKIMSKODNCVHDLQOHKEIYTIKWSPTGPGTNNFNANMLASAF 487
QY 421 DSTVRLWDVDRGICHTLTKHQBVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 472
DB 488 DSTVRLWDVDRGICHTLTKHQBVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 539

RESULT 7
ABB60376
ID ABB60376 standard; protein; 700 AA.
XX
AC ABB60376;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7920.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;
WPI; 2001-656860/75.
N-PSDB; ABL04479.
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions.
XX Disclosure; SEQ ID NO 7920; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (AB857737
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 700 AA;

Query Match 63.1%; Score 1715.5; DB 4; Length 700;
Best Local Similarity 52.9%; Pred. No. 1.3e-151;
Matches 349; Conservative 49; Mismatches 71; Indels 191; Gaps 8;

QY 1 MSISDEVNPLVRYLOESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSFSDEVNPLVRYLOESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
QY 61 EARSINEDGTLFDRPIESLIDAVMPDV-----VQTRQ----- 96
DB 61 EVESVSGEDGEV--ARPIEGLSLIDAVMPDVKPLKPIVKTPEKPGAVDSSAPAGNQNN 118
QY 97 -----QAYRDKLA-----QQQAAAAAASQSGSAKQNGENT 130
DB 119 NAKPEIKETPGTVAGAGGKNKTAGSTGTSTPTDQASVDSGNNANNAGTYAGNG 178
QY 131 ANGBE----- 135
DB 179 AGGNQASTGSGNSTPAGGDLAAPGASQKKSQNSNEAGSSSGNAGNANATSTDDAAS 238
QY 136 --NG-----AHTIANNHTDM----- 148
DB 239 TSTNGNSTSSVEQPTSGLTTPAGGTVSTNPNDAASGGASTATGSKAPGAVTIRVGAQ 298
QY 149 -----NEVDGDVEIPPNKAVLRGHESEV 172
DB 299 GNNVQSGSSNAQSSAPSGTISSTSGGAGTAPALVPMIDENIEIPSKARVLRGHESEV 358
QY 173 FICAWNPSDLLASGSDSTARINWLSNSTSGTQVLVLRHCIRREGQDVPSNKDVTSLD 232
DB 359 FICAWNPSDLLASGSDSTARINWLSNSTSGTQVLVLRHCIRREGQDVPSNKDVTSLD 417
QY 233 WNSGTLATGSYDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGFIILSAGVDKTTII 292
DB 418 WNCDSLLATGSYDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGFIILSAGVDKTTII 477
QY 293 WDAHTGEAKQOFPFHSAAPALDVWDQSNNTFASCSTDMCHVCKLQODRPKTFQHTNEV 352
DB 478 WDAHTGEAKQOFPFHSAAPALDVWDQSNNTFASCSTDMCHVCKLQODRPKTFQHTNEV 537
QY 353 NAIKWDPTGNLLASCDSDMTLKIMSKODNCVHDLQOHKEIYTIKWSPTGPGTNNFNAN 412
DB 538 NAIKWCPOGQLLASCDSDMTLKIMSKODNCVHDLQOHKEIYTIKWSPTGPGTNNFNAN 597
QY 413 LMLASAFDSTVRLWDVDRGICHTLTKHQBVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 472
DB 598 LILASAFDSTVRLWDVDRGICHTLTKHQBVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 657

RESULT 8
AAO04385
ID AAO04385 standard; protein; 208 AA.
XX
AC AAO04385;
XX
DT 06-NOV-2001 (first entry)


```
XX DE Human polypeptide SEQ ID NO 18277.
XX DE
XX DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US004927.
XX PR 28-FEB-2000; 2000US-00515126.
XX PR 19-MAY-2000; 2000US-00577409.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-514838/56.
XX DR N-PSDB; AA184316.
XX PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX PT and treating e.g. leukemia, inflammation and immune disorders.
XX PS Claim 20; SEQ ID NO 18277; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AA179941-AA193841) and
XX CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation. Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 208 AA;
XX
XX Query Match 41.8%; Score 1137; DB 4; Length 208;
XX Best Local Similarity 99.5%; Pred. No. 4.3e-98;
XX Matches 206; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 213 HCIREGGQVPSNKDVTSLDWNSGTLATGSDGPFARITWKGDLASTLGGHKGPIPAL 272
XX Db 1 HCIREGGQVPSNKDVTSLDWNSGTLATGSDGPFARITWKGDLASTLGGHKGPIPAL 60
XX
XX QY 273 KNNKGNFSLAGVDKTTIIWDAHTGEAKQOFPFHSAPALVDWQSNNTFASCSGDMCIH 332
XX Db 61 KNNKGNFSLAGVDKTTIIWDAHTGEAKQOFPFHSAPALVDWQSNNTFASCSGDMCIH 120
XX
XX QY 333 VCKLGDRIKTFQGTNEVNAIKDPTGNLLASGDDMTLKIWSMKQDNCVHDJQOENK 392
XX Db 121 VCKLGDRIKTFQGTNEVNAIKDPTGNLLASGDDMTLKIWSMKQDNCVHDJQOENK 180
XX
XX QY 393 EYTIKWSPTGPTNNPNANMLASAS 419
XX Db 181 EYTIKWSPTGPTNNPNANMLASAS 207
XX
XX RESULT 9
XX ABG21350
XX ID ABG21350 standard; protein; 167 AA.
XX AC ABG21350;
```

```
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #21341.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS85537.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 51709; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 167 AA;
XX
XX Query Match 30.1%; Score 819; DB 4; Length 167;
XX Best Local Similarity 86.2%; Pred. No. 2.3e-68;
XX Matches 144; Conservative 13; Mismatches 10; Indels 0; Gaps 0;
XX
XX QY 250 RIWTKDGNLASTLGOHKGPIFALKWNNKGNFISAGVDKTTIIWDAHTGEAKQOFPFHS 309
XX Db 1 RIWTKDGNLASTLGOHKGPIFALKWNNKGNFISAGVDKTTIIWDAHTGEAKQOFPFHS 60
XX
XX QY 310 PALDQVDMQSNNTFASCSGDMCIHVCKLGDRIKTFQGTNEVNAIKWDTGNLLASCS 369
XX Db 61 PALDQVDMQSNNTFASCSGDMCIHVCKLGDRIKTFQGTNEVNAIKWDTGNLLASCS 120
XX
XX QY 370 DMTLKIWSMKQDNCVHDJQOENKEIYTIKWSPTGPTNNPNANMLA 416
XX Db 121 DMTLKIWSMKQDNCVHDJQOENKEIYTIKWSPTGPTNNPNANMLA 167
```

QY	220	QDVPSNDKVTSLDWNSSGTTLLGSGVDGFGFARIWTKDGNLASTLGOHKGPTIFALKWNNKGN	270
DB	215	SSGKTTNQVTCIAWSHDGNSIVTGVENGELRWKTKGALLNVNFRAPIVSVKWNKDTG	274
QY	280	PILSAGVDKTTIIDAHTGEAKQQFPF-----HSAP-----ALDVIDMOSNNTFAS	324
DB	275	HIISMDVENVTILNWNISGTVMQHFELKETGGSSINAENHSBGSLGVDVVEWDDDKFVI	334
QY	335	CSTDMCIVCKLGDRIKTKFOGHTNEVNAIKWDPGTCNLLASCSDDMTLKINSMKQDNCV	384
DB	335	PCPKGAIFVYQITEKTKTGKLIHGHPISVLEFENDTKLLLSASDDGTURIHGGNGNSQ	394
QY	385	HDLQCHNKIEIYTIKWSPTGCTNNPNANMLASASFDSTVRLMDVDVRGICHTLTKHQEP	444
DB	395	NCFVCHSQSIVSASFW-----GDDKVISCSMDGSVRLSLKQNTLLALSIVDGVF	444
QY	445	VYSVAFSGDGRYLAGSGFDKCVHIWNTQVCLHYINGQVLLNLG-RSICLYTLPHHLVVP	503
DB	445	IFAGRISQDGKAYAFPMGQVNVYD-----LKKLNGSKRSRSLYGNRDLNPLP-----IP	495
QY	504	LVA 506	
DB	496	LVA 498	
RESULT	11		
ADAL3321			
ID	ADAL3321	standard; protein; 414 AA.	
XX	ADAL3321;		
XX	AC		
XX	DT		
XX	DT		
XX	DT		
DE	06-NOV-2003	(first entry)	
XX	Human intracellular signalling molecule INTSIG-3, SEQ ID NO:3.		
XX	Human; intracellular signalling molecule; INTSIG;		
XX	cell proliferative disorder; cancer; atherosclerosis;		
KW	autoimmune disorder; inflammatory disorder; infection;		
KW	neurological disorder; developmental disorder; endocrine disorder;		
KW	cytostatic; antiarteriosclerotic; nootropic; neuroprotective;		
KW	cerebroprotective; immunosuppressive; antiinflammatory; gene therapy.		
OS	Homo sapiens.		
XX	WO2003031568-A2.		
XX	PN		
XX	PN		
PD	17-APR-2003.		
XX	16-AUG-2002; 2002WO-US026322.		
XX	17-AUG-2001; 2001US-0313245P.		
PR	24-AUG-2001; 2001US-0314751P.		
PR	31-AUG-2001; 2001US-0316752P.		
PR	31-AUG-2001; 2001US-0316847P.		
PR	14-SEP-2001; 2001US-0322188P.		
PR	28-SEP-2001; 2001US-0326390P.		
PR	12-OCT-2001; 2001US-0328952P.		
PR	19-OCT-2001; 2001US-0345468P.		
PR	12-APR-2002; 2002US-0372499P.		
XX	(INCYTE) INCYTE GENOMICS INC.		
XX	PA		
XX	Yue H, Lu DAM, Swarnakar A, Tang YT, Griffin JA, Emerling BM;		
PI	Forsythe TJ, Yao WG, Ramkumar J, Richardson TW, Becha SD, Lee EA;		
PI	Warren BA, Lehr-Mason PM, Baughn MR, Li JX, Duggan BM, Gietzen KJ;		
PI	Lal PG, Borowsky ML, Ison CH, Thangavelu K, Xu Y, Lee S;		
PI	Elllott VS, Sprague WW, Azimzai Y, Hafalia AJA, Ding L, Nguyen DB;		
PI	Honchell CD, Luo W, Chawla NK, Marquis JP, Jackson JL, Tran UK;		
XX	WPI; 2003-393436/37.		
DR	DR N-PSDB; ADA13366.		
XX	PT		
XX	New human intracellular signaling molecules (INTSIG)), useful for		

PT diagnosing, treating and preventing diseases or conditions associated
 PT with the aberrant INTSIG expression e.g. cancer, diabetes, epilepsy, or
 XX infections.

Claim 1; Page 225-226; 367pp; English.

XX The invention relates to 45 human intracellular signalling molecules
 CC (INTSIG), INTSIG-1 to INTSIG-45, and the cDNAs encoding them (ADAL319-
 CC ADAL3408). The invention also encompasses expression constructs, host
 CC cells and transgenic organisms comprising an INTSIG nucleic acid sequence
 CC ; the recombinant preparation of a INTSIG; an antibody against a INTSIG;
 CC methods of detection of INTSIG proteins or nucleic acids; a micro-array
 CC containing INTSIG nucleic acids; methods of screening compounds for their
 CC ability to modulate INTSIG activity or expression; and pharmaceutical
 CC compositions comprising an INTSIG protein, an INTSIG antibody, an INTSIG
 CC agonist or INTSIG antagonist. The INTSIG proteins, nucleic acids or
 CC compositions comprising them are useful in diagnosing, treating or
 CC preventing a variety of disorders, including cell proliferative disorders
 CC (e.g., cancer or atherosclerosis); autoimmune/inflammatory disorders
 CC (e.g., AIDS, allergies, anaemia, asthma, bronchitis, gout, multiple
 CC sclerosis, osteoarthritis, rheumatoid arthritis, or systemic lupus
 CC erythematosus); neurological disorders (e.g., epilepsy, stroke,
 CC Alzheimer's disease, dementia, or Parkinson's disease); viral, bacterial,
 CC fungal, parasitic, protozoan or helminthic infections; neurological
 CC disorders (e.g., epilepsy, stroke, Alzheimer's disease, dementia, or
 CC Parkinson's disease); developmental disorders (e.g., achondroplastic
 CC dwarfism, Cushing's disease, hypothyroidism or hydrocephalus); endocrine
 CC disorders (e.g., disorders of the hypothalamus or pituitary gland or
 CC diabetes); reproductive disorders; or vesicle-trafficking disorders
 CC (e.g., cystic fibrosis or hypercholesterolaemia). The present sequence
 CC represents an INTSIG protein of the invention.

XX Sequence 414 AA;

Query Match 13.8%; Score 374.5; DB 6; Length 414;

Best Local Similarity 28.3%; Pred. No. 5.7e-26;
 Matches 99; Conservative 49; Mismatches 137; Indels 65; Gaps 10;

QY 129 NTANGENGAAHTIANNHTDMEVDGVEIPPKAVVLRGHSEVFICAW-NPVSDLLASG 187
 DB 120 DTASGEELN-----TLEGRNVVVAIAENPEYGDKIATG 153
 QY 188 SGDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSKDVTSLDWNSEGTLLATGSDYG 247
 DB 154 SPDKTKLMSVETG-----CYHTF-----GHTAEIVCLSFNPOSTLVATGSMGT 199
 QY 248 FARINTKDCNLASTLQGHKGPFPALKWKKGNFILLSAGVDKTTIWDADHTGEAKQFPFH 307
 DB 200 TAKLWDIQNGEELTLRGHSAEILISFNTSGRIITGSPDHTVVVWDADTGRKNVILICH 259
 QY 308 SA--PALDYDWSNNTFASCS---TDMCIHVCKL---GQDRPIKTFQGHNEVNAIKWDP 359
 DB 260 CAEIGSASFN-----DCSLILTGSMDTKCKLWDATNGKCVATLTGHDDEILDSCFDY 312
 QY 360 TGNLLASCSDDMTLKIWSKQDNCVHDLOQHKEIYTIKWSPTGCTNNPNANLMLASAS 419
 DB 313 TGKLTASADGTARIFSAATKCAKLEHGEIGESKISF-----NPQGNHILLTGS 364
 QY 420 FDSVRLWVDVRGICHTHTTKHQPVSVAFPDGRYLASGSDKCVHIW 469
 DB 365 -DKTARIWDAQTGCQLQVLEHTDEIFSCAFNYKGNVITGSKDNTCRIW 413

RESULT 12

ID ABB62260
 ID ABB62260 standard; protein; 411 AA.

XX ABB62260;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 13572.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL06363.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.

XX Disclosure; SEQ ID NO 13572; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABS72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 411 AA;

Query Match 13.6%; Score 370.5; DB 4; Length 411;

Best Local Similarity 22.4%; Pred. No. 1.3e-25;
 Matches 115; Conservative 77; Mismatches 158; Indels 163; Gaps 14;

QY 6 DEVNFLVRYLQESGFSGHSFTFIKSHIS---QSNINGALVPPAALISITIKGLQVVEA 62
 DB 11 EELNQAADYLGSGNYADSLSTFEKADLSTEVEKFGGLLEKKTWSVIRLQKKWLEA 70
 QY 63 EVSINEBGTLPDGRPIESLSLIDAVMPDVVQTQQAYRDKLAQQQAAAAAASQQG 122
 DB 71 KLTEAE-----KEVIE----- 81
 QY 123 SAKGENTANGEENGAAHTIANNHTDMEVDGVEIPPKAVVLRGHSEVFICAWNPSVD 182
 DB 82 ---GATKKKRTPEW-----TPRPEK-FSLTGHRASTRVIFHFIFA 121
 QY 183 LLASGSDSTARIWNLS-----ENSTSGTQLVLRHCIREGGQDVPSKDVTSLDWNSEGT 238
 DB 122 LMVSASEDATIRIWDFTGTEYERSLKGTDSV-----QDVA-----FDAQGX 163
 QY 239 LLATGSDYDGFARIW--TKDGNLSTLQGHKGPFPALKWKKGNFILLSAGVDKTTIWDAD 296
 DB 164 LLASCSADLSIKLWDFQOSTECIKTMHGHNDHNVSVAFPAGDYVLTSASDRTIKMEVA 223
 QY 297 TGEAKQOFPFHSAFALPDVDMQSNNTFASCSDTMCIHVCKLQDRPIKTFQGHNEVNAIK 356
 DB 224 TG-----YC-----VKTYTGHREWVWVR 242
 QY 357 WDPGTNLLASCSDDMTLKIWSKQDNCVHDLOQHKEIYTIKWSP-----TGPGTNN 408
 DB 243 VHIEGTFATCSNDQTIIRVMTLNSKCKVELRDEHTVECIANAPEAAASAINAEAGADN 302
 QY 409 PNAN---LMLASASFDSVRLWVDVRGICHTHTTKHQPVSVAFPDGRYLASGSDK 465

Db 303 KKGHGQPLASGRDKTIRINDVSVGLCLLTLGSHDNVVRGLAFHPGKYLVSASDDKT 362
 QY 466 VHIWNTQVCLHYLNGQVLLNGLRSICLYTLPHH 498
 Db 363 IRVW-----DLNKRCKMITYAH 380

RESULT 13
 AAB68284
 XX AAB68284 standard; protein; 485 AA.
 AC AAB68284;
 XX
 DT 09-JUL-2001 (first entry)
 XX
 DE Amino acid sequence of a human Tzap gene cDNA clone.
 XX
 KW Tzap; T cell activation; immune response; transplant rejection;
 KW bone marrow transplantation; rheumatoid arthritis; lupus erythematosus;
 KW multiple sclerosis; encephalitis; vasculitis; diabetes mellitus;
 KW pancreatitis; gastritis; thyroiditis; malignant disorder; asthma;
 KW lepramatosis; gastritis; skin tumour; adrenal tumour; lung tumour;
 KW wound healing; growth disorder; inflammatory disease; infectious disease.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 379
 FT /note= "unspecified residue encoded by NTC"
 FT
 XX W0200132614-A2.
 XX
 PD 10-MAY-2001.
 XX
 XX 30-OCT-2000; 2000WO-EP010670.
 XX
 XX 01-NOV-1999; 99US-0162675P.
 PR 25-FEB-2000; 2000US-0185016P.
 XX
 XX (UTKU/) UTKU N.
 PA
 XX Utku N;
 PI
 XX WPI; 2001-316410/33.
 DR N-PSDB; AAF85075.
 DR
 XX Novel polynucleotide encoding Tzap protein involved in T cell activation,
 PT useful for diagnosing and treating diseases involving T cell activation,
 PT for treating organ transplantation rejection, rheumatoid arthritis.
 XX
 XX Example; Page 56-57; 68pp; English.
 PS
 XX The present sequence represents a Tzap protein. The Tzap gene is involved
 CC in T cell activation, and in the modulation of immune responses. Tzap
 CC polynucleotides and polypeptides are useful for diagnosing or treating
 CC acute and chronic diseases involving T cell activation and Th1 and Th2
 CC immune response, for the treatment of acute and chronic rejection of allo
 CC - and xeno organ transplants and bone marrow transplantation, for the
 CC treatment of rheumatoid arthritis, lupus erythematosus, multiple
 CC sclerosis, encephalitis, vasculitis, diabetes mellitus,
 CC pancreatitis, gastritis, thyroiditis, for the treatment of malignant
 CC disorders of T, B or NK cells, for the treatment of asthma, lepramatosis,
 CC Helicobacter pylori associated gastritis or for the treatment of skin
 CC tumours, adrenal tumours or lung tumours, wound healing, growth disorders
 CC inflammatory and/or infectious diseases
 CC
 XX Sequence 485 AA;
 SQ

Query Match 13.5%; Score 367.5; DB 4; Length 485;
 Best Local Similarity 26.8%; Pred. No. 3.4e-25;
 Matches 103; Conservative 47; Mismatches 145; Indels 89; Gaps 9;

QY 165 LRGHSEVFICAMNPVSDLLASGSDSTARIWNLSEN-----STSGSTQVLVLRHCIREGGQ 220
 Db 110 LEHSEAVISVAFSPGKYLASGSDTTVRFWDLSTETPHFTCKGHRHWLVSISWSFDGK 169
 QY 221 DVPS-----NKDVTSLDW-----NSEGTLATGSYDGFAR 250
 Db 170 KLASGCKNGOILLWDPSTGKQVGRITLAGHKSKITGLSWEPLHANPECRYVASSSKDGSVR 229
 QY 251 IW-TKQGNLASTLQOHKGPFPALKNKKGNFILLSAGVDKTIITWDAHTGEAKQPFPH-- 307
 Db 230 IMDTTAGRCRILITGHTQSVTCLRWGGDG-LLYSASQDRTIKVRAHDGVLCRTLQCHGH 288
 QY 308 --SAPALDVM-----QSNNTFASGS 326
 Db 289 WWTMALSTDYALRTGAFEPABASVNPQDLQSGIQLKRALSRYNLVRGQGERLVSQS 348
 QY 327 TDMCIHVCKLQGD-RPIKTFQGHTEVNAIKWPTGNLASCDDMTLKWSVKQDNCVH 385
 Db 349 DDFTLFLWSPAEDKKPLTRMTGHQALINQVXFPDSRIVASAFDKSIKLMWDGRTGKYLA 408
 QY 386 DLQOHKEIYTIKWSPTGPGTNNPNANMLASAFDSTVRLVDVDRGICITHTLTKEQEPV 445
 Db 409 SLRGHVAAVQIAWSAD-----SRLVSGSSDSTLKVDVKAKQKLANDLPGRHADEV 459
 QY 446 YSAFSPDGRYLASGSPDKCVHIW 469
 Db 460 YAVDWSPDGQVAVSGGKDKCLRIW 483

RESULT 14
 AAB68516
 ID AAB68516 standard; protein; 485 AA.
 XX
 AC AAB68516;
 XX
 DT 19-APR-2001 (first entry)
 XX
 DE Human GTP-binding associated protein #16.
 XX
 KW Human; Guanosine triphosphate binding associated protein; GTP; GRAP;
 KW inflammation; AIDS; Addison's disease; anaemia; arteriosclerosis; asthma;
 KW autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;
 KW osteoporosis; psoriasis.
 XX
 OS Homo sapiens.
 XX
 XX W0200105970-A2.
 PD 25-JAN-2001.
 XX
 PF 19-JUL-2000; 2000WO-US019698.
 XX
 PR 19-JUL-1999; 99US-0144595P.
 PR 23-AUG-1999; 99US-0150460P.
 PR 15-OCT-1999; 99US-0159849P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 PI Yue H, Tang YT, Bandman O, Hillman JL, Lal P, Au-Young J;
 PI Reddy R, Yang J, Baughn MR, Lu DAM, Azimzai Y, Patterson C;
 XX
 DR WPI; 2001-091972/10.
 DR N-PSDB; AAF58316.
 XX
 PT New guanosine triphosphate-binding associated proteins (GBAP) and their
 PT encoding nucleic acids, useful for treating and/or diagnosing diseases
 PT associated with GBAP expression, such as cancer, diabetes and asthma.
 XX
 PS Claim 1; Page 133-134; 233pp; English.
 XX
 CC The present invention relates to novel human guanosine triphosphate (GTP)
 CC -binding associated proteins (GBAPs; AAB68501-AAB68566) and their coding
 CC sequences (AAF58301-AAF58366). The proteins and coding sequences of the

CC present invention are useful for treating a variety of disorders
CC including inflammation, AIDS, Addison's disease, anaemia,
CC arteriosclerosis, asthma, autoimmune disorders, Grave's disease,
CC hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and
CC psoriasis

XX Sequence 485 AA;

Query Match 13.5%; Score 366.5; DB 4; Length 485;
Best Local Similarity 26.8%; Pred. No. 4.2e-25;
Matches 103; Conservative 47; Mismatches 145; Indels 89; Gaps 9;

QY 165 LRGHSEVFICAWNPVSDLLASGSGDSTARIWNLSN-----STSGTQLVLRHCIREGQ 220
DB 110 LEHSEAVISVAFSPGKYLASGSDTIVRFDLSTPHFTCKGHRHWLWSISWSPDGK 169
QY 221 DVPS-----NKDVTSLDW-----NSEGTLATGSDYDGFAR 250
DB 170 KLASGCKNGQILLWDPSTGKQVGRTLAGHSKWITGLSWEPLANPECRYVASSSKDGSVR 229
QY 251 IW-TKDGNIASLTGQHKGPFAKWNKGNFILSAGVDKTTIWDATGEAKQOFPFH-- 307
DB 230 IWDTTAGRCERILTGHTQSVTCLRWGGDG-LLYSASQDRTIKVRAHDGVLCTLQGHGH 288
QY 308 --SAPALDWDW-----QSNNTFASCS 326
DB 289 WNTMALSTDYALRTGAFEPASVNPQDLQGSJQELKERALSRYNLVRGQPERLVSGS 348
QY 327 TDMCIHVCKLQGD-RPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVH 385
DB 349 DDFTFLWSPAEKDKPLTRMTGHOALINQVLFSPDSRIVASASFDKSIKLDWGRGKYLA 408
QY 386 DLOOHKEIYTIKWSPTGPTGNPNANMLASASFDSTVRLWDVDRGICHTLTKHOEPV 445
DB 409 SLRGHVAAYQIAWSAD-----SRLVSGSSDSTLKWDVKAQKLANDLPCHADEV 459
QY 446 YSVAFSPDGRYLASGSDKCVHIW 469
DB 460 YAVDWSPDGQRVASGKDKCLRIW 483

RESULT 15

AAB92844
ID AAB92844 standard; protein; 485 AA.

XX AAB92844;

XX 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:11400.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.

XX Claim 8; SEQ ID NO 11400; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC of an oligonucleotide comprising at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

XX Sequence 485 AA;

Query Match 13.5%; Score 366.5; DB 4; Length 485;
Best Local Similarity 26.8%; Pred. No. 4.2e-25;
Matches 103; Conservative 47; Mismatches 145; Indels 89; Gaps 9;

QY 165 LRGHSEVFICAWNPVSDLLASGSGDSTARIWNLSN-----STSGTQLVLRHCIREGQ 220
DB 110 LEHSEAVISVAFSPGKYLASGSDTIVRFDLSTPHFTCKGHRHWLWSISWSPDGK 169
QY 221 DVPS-----NKDVTSLDW-----NSEGTLATGSDYDGFAR 250
DB 170 KLASGCKNGQILLWDPSTGKQVGRTLAGHSKWITGLSWEPLANPECRYVASSSKDGSVR 229
QY 251 IW-TKDGNIASLTGQHKGPFAKWNKGNFILSAGVDKTTIWDATGEAKQOFPFH-- 307
DB 230 IWDTTAGRCERILTGHTQSVTCLRWGGDG-LLYSASQDRTIKVRAHDGVLCTLQGHGH 288
QY 308 --SAPALDWDW-----QSNNTFASCS 326
DB 289 WNTMALSTDYALRTGAFEPASVNPQDLQGSJQELKERALSRYNLVRGQPERLVSGS 348
QY 327 TDMCIHVCKLQGD-RPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVH 385
DB 349 DDFTFLWSPAEKDKPLTRMTGHOALINQVLFSPDSRIVASASFDKSIKLDWGRGKYLA 408
QY 386 DLOOHKEIYTIKWSPTGPTGNPNANMLASASFDSTVRLWDVDRGICHTLTKHOEPV 445
DB 409 SLRGHVAAYQIAWSAD-----SRLVSGSSDSTLKWDVKAQKLANDLPCHADEV 459

Search completed: August 9, 2004, 16:45:26
Job time : 50 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: August 9, 2004, 16:41:43 ; Search time 13.333 Seconds
(without alignments)
3708.183 Million cell updates/sec
Title: US-09-987-701-4
Perfect score: 2718
Sequence: 1 MSISDEVNFLVRYLQESG.....LPHLWVPLVALIELLVK 514
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	475	17.5	1526	2 AC2239	WD-40 repeat prote
2	472	17.4	535	2 S48268	probable membrane
3	458.5	16.9	1258	2 AI2155	WD-repeat protein
4	451.5	16.6	1683	2 AF2071	WD-40 repeat prote
5	444.5	16.4	1227	2 AE1810	WD-40 repeat prote
6	429.5	15.8	1189	2 AI2493	WD-repeat protein
7	426	15.7	1708	2 AE1866	WD-40 repeat prote
8	420.5	15.5	1356	2 T18521	beta transducin-li
9	408	15.0	1711	2 AD1842	WD-40 repeat prote
10	404.5	14.9	934	2 AG1889	WD-40 repeat prote
11	395	14.5	1747	2 AC1842	WD-40 repeat prote
12	394.5	14.5	1693	2 S76086	WD-40 repeat prote
13	382.5	14.1	564	2 T40883	beta transducin-li
14	381	14.0	1351	2 AE2410	WD repeat protein
15	369	13.6	676	2 AH2195	WD-repeat protein
16	368.5	13.6	304	2 AG1837	hypothetical prote
17	366	13.5	502	2 T41148	WD-40 repeat prote
18	365	13.4	1189	2 AH2154	WD-repeat protein
19	354	13.0	265	2 AF1890	WD-repeat protein
20	349.5	12.9	410	2 S48052	platelet-activatin
21	344.5	12.7	559	2 AD2202	hypothetical prote
22	342	12.6	409	2 S36113	LIS-1 protein - hu
23	341.5	12.6	677	2 AE1861	serine/threonine k
24	339.5	12.5	786	2 AG2375	WD-40 repeat-prote
25	337	12.4	515	2 S19487	hypothetical prote
26	332	12.2	323	2 T02617	hypothetical prote
27	331	12.2	437	2 S05357	hypothetical prote
28	326.5	12.0	787	2 T00798	hypothetical prote
29	324	11.9	342	2 AE2490	WD-repeat protein

WD-repeat protein
probable WD-repeat
PF20 protein, micr
hypothetical prote
probable sulfur me
WD-40 repeat prote
hypothetical prote,
transcription init
TATA box-binding p
WD-40 repeat regul
hypothetical prote
hypothetical prote
MET30 protein - ye
beta transducin-li
hypothetical prote
hypothetical WD-re

ALIGNMENTS

RESULT 1
AC2239
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2239
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2239
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1526 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA000019; PID:gl7132599; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3466

Query Match 17.5%; Score 475; DB 2; Length 1526;
Best Local Similarity 31.0%; Pred. No. 7.1e-26;
Matches 119; Conservative 63; Mismatches 138; Indels 64; Gaps 10;
QY 136 NGARTIANNTDMVEVDGVEIPENKAV-VLRGHESEVFICAWNPVSDLLASGSDSTAR 194
DB 1127 NGV-TLANGSDQIVRLWD--ISSKCLYTLQGHNNWNAVAFSPDGATLASGGDQTVR 1183
QY 195 IWNLSNS---TSGSTQLVLRHCIREGGQDVFPSNKDVTSLDNWSEGTLLATGSDGFAR 250
DB 1184 LWDISSKCLYILQGHTSW-----VNSVFNPDGSLASGSDQTVR 1225
QY 251 IW-TKDNGLASTLQGHKGPIFALKNKKNKGNFILSAGVDKTIIWDAHTGAQKQFPFHS 309
DB 1226 LWEINSSKCLCTFGHTSWNSVFNPDGSMASGSDKTVRLWDISSKCLHTFQGH 1284
QY 310 PALDWDVQSNNTF-----ASCSTDMCIHVCKLQGRPIKTFQGHNEVNAIKWDPTGN 362
DB 1285 -----NNVNSVAFNPDGSMLASGSDQTVRLWISSKCLHTFQGHISWVSVPDGT 1339
QY 363 LLASCSDDMTLKWSMKQDNCVHDLQHNKEIYTIKWSPTGPTGNTNPNANMLASAFDS 422
DB 1340 MLASGSDQTVRLWISSSGECLYTLFGHTNWGVSIFSPDG-----AILASGSDQ 1390
QY 423 TVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYLAGSPDKCVHIWNTQVCLHYLNGQV 482
DB 1391 TVRLWISSKCLYTLQCHNNWVGSIWVSPDG-----AILASGSDQTVRLW 1437
QY 483 LLNLGRSICLYTLPHHLVWIPPLVA 506
DB 1438 --NISSGECLYTLHGINSVRSVA 1459

Db 445 IFAGRISQDGQKYAVAFMDQGVNVD-----LKKLNKSRSLYGNRDGILNPLP-----IP 495

Qy 504 LVA 506

Db 496 LYA 498

RESULT 3

AI2155

WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AF2071

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, R.; Kaneko, T.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.; Nakazaki, N.; 205-213, 2001

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AF2071

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1258 <KUR>

A:Cross-references: GB:BA000019; PIDN:BAE74499.1; PID:gl7131893; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr2800

Query Match 15.9%; Score 458.5; DB 2; Length 1258;

Best Local Similarity 32.0%; Pred. No. 8.5e-25;

Matches 105; Conservative 43; Mismatches 101; Indels 79; Gaps 8;

Qy 165 LRHSESEVFICAWNPVSDLLASGSDSTARIMLSENSTSGTQVLVLRHCIREGGQDVP 224

Db 974 LEHTDFYIGAFSPDSQTLASASTDSVRLNI---STGQCQILLEH----- 1019

Qy 225 NKDVTSLDW-----NSEGTLLATSYDGFARIWT-XDGNLSTLGOHQKGPFIALKWKK 277

Db 1020 -----TDWYAVVFHQKIIATGSADCTVKLWNISTGQCLKTLSEHSDKILGMAWSPD 1073

Qy 278 GNFLSAGVDKTTIWDHAHTGEAKQPFHSAPALPDVDWQSNNTFASCSTDCMCHVCKLG 337

Db 1074 GQLLASASADQSVRLWD-----CCTGRVCVI----- 1099

Qy 338 QDRPIKTFQGHTEVNAIKWDPCTGNLLASGDDMTLKIWSKQDNCVHDLQHNKEIYTI 397

Db 1100 -----LRHSESEVFICAWNPVSDLLASGSDSTARIMLSENSTSGTQVLVLRHCIREGGQDVP 1152

Qy 398 KWSPTGPTNNPNANMLASASFDSTVRLWDVDRGICIHILTKHQBEPVYSVAFSPDGRYL 457

Db 1153 AFSPDGK-----LLASASHDQTVRIWDVNTGKCHICIGTHLVSSVAFSPDGEV 1203

Qy 458 ASGSFQKCVHIWN--TOVCLHLYNGQVL 483

Db 1204 ASGSQDQTVRIWNVKTEGCLQLAKRL 1231

RESULT 4

AF2071

WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AF2071

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, R.; Kaneko, T.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.; Nakazaki, N.; 205-213, 2001

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AF2071

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1683 <KUR>

Db 445 IFAGRISQDGQKYAVAFMDQGVNVD-----LKKLNKSRSLYGNRDGILNPLP-----IP 495

Qy 504 LVA 506

Db 496 LYA 498

RESULT 2

S48268

probable membrane protein YBR103w - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBR0832

C:Species: Saccharomyces cerevisiae

C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C:Accession: S48268; S45971; S44683

R:Wannhaup, G.; Stucka, R.; Ehmele, S.; Vetter, I.; Feldmann, H.

Yeast 10, 1363-1381, 1994

A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.

A:Reference number: S48255; MUID:95208357; PMID:7900426

A:Accession: S48268

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-535 <WAN>

A:Cross-references: EMBL:X78993; NID:5476045; PIDN:CAAS5606.1; PID:9476059

R:Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45971

A:Accession: S45971

A:Molecule type: DNA

A:Residues: 1-535 <FE2>

A:Cross-references: EMBL:Z35972; MIPS:YBR103w

C:Genetics:

A:Gene: SGD:SIF2

A:Cross-references: SGD:S0000307

A:Map position: 2R

C:Map family: unassigned WD repeat proteins; WD repeat homology

C:Keywords: transmembrane protein

F:355-388/Domain: WD repeat homology <WD1>

F:397-429/Domain: WD repeat homology <WD2>

F:434-450/Domain: transmembrane #status predicted <TM>

Query Match 17.4%; Score 472; DB 2; Length 535;

Best Local Similarity 25.8%; Pred. No. 2.8e-26;

Matches 140; Conservative 91; Mismatches 230; Indels 82; Gaps 15;

Qy 1 MSITSEELNYLWYCYCEMGHEVSALALQDETRLFDEKYEKHI-----PLGTL 50

Db 50 ISIIQKGLQVHAEVSNEDG---TLFDGRIPLSLIDAVMPDVVQTSQAVDQLAQ 106

Db 51 VNLVQRGILYTESLMDVSKGDISALNEHLSDFNLVQALQID-----KEAFPEI 101

Qy 107 QAAAAAASQGSQAKNGENTANGENGATIANNHDTMMVEVDGVEIPENKAVVLR 166

Db 102 SSEGRFLETSNENKAGEDGASIVRETOEDDTSIDSSD--DLDGFKVI--LKEIV-- 155

Qy 167 GHSESEVFICAWNPVSD--LLASGSDSTARIMLSENSTSGTQ---LVLRHCTREGG 219

Db 156 -KLDNIIVSTWNPDESILAYGEKNSVARLARIIVETDQEGKTKWKTITAELRHPFALSA 214

Qy 220 QDVPSNKDVTSLDWNSGTLTATGSYDGFARIWT-KDGNLSTLGOHQKGPFIALKWKKGN 279

Db 215 SSGKTTNQVTLAWSHGNSIVIVGENGELNKNKTGALLNLVNFHRAPIVSKWKKOGT 274

Qy 280 FILSAGVDKTTIWDHAHTGEAKQPF-----HSAP-----ALVDWQSNNTFAS 324

Db 275 HIISMDVENVTILNVIISGTVMOHFELKETGGSSINAENHSGDGLGVDEWVDDDKFVI 334

Qy 325 CSTDWCHVCKLGDRPIKTFQGHTEVNAIKWDPCTGNLLASGDDMTLKIWSKQDNCV 384

Db 335 PGPKGALFVYQITETKTKLIGHGPIVLEFNDTNKLLSASDDGTLRIHGGGNSQ 394

Qy 385 HDLQHNKEIYTIKWSPTGPTNNPNANMLASASFDSTVRLWDVDRGICIHILTKHQB 444

Db 395 NSFYGHQSIVSAGW-----GDDKVISCMDSGVRLWSLKNITLLALSIVDGV 444

Qy 445 VYSVAFPDGGRYLAGSGFDKCVHIWNTQVCLHYLNGQVLNLG-RSICLYTLPHHLVVP 503

QY 118 ASQQSAK-----NQE--NTANGEENGAAHTIANNHTDMEYDG-----DV 150

DD T56 WWWALAFSTDT-----DAXWASGSIDHIVATWDSGGCCCTDQGHFSSVAVATSC

```
Qy      453 DCRYLASGSPDKCVHWN--TQVCL-----HYL---NGQV---LLN 485
       |||:::|
Db      1002 DKKTLFSSGEKLVKQNDVEYCYCLOTWEADNRVMAVAVSRDNYQLATGGDSVVRLWD 1061
```

Qy	486	LGRSICLYTLPHH	498
		: : :	
Dd	1062	IGKGVCVRTFSGH	1074

RESULT 7

AE1866
WD-400 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AE1866
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Tanaka, K.; Shimizu, T.; Terauchi, T.; Inaba, H.; Miyajima, N.; Murakami, K.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE1866
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1708 <XUR>
A;cross-references: GB1BA00001.9; PIDN:BA872436.1; PID:gw17129823; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all0478

```

Query Match      15.7%; Score 426; DB 2; Length 1708;
Best Local Similarity 32.4%; Pred. No. 2.9e-22;
Matches 110; Conservative 61; Mismatches 130; Indels 42; Gaps 11;

QY 132 NGEAAGHTIANNHDTMVEGDGVEIPPNKAVV---LRGHSESVFICAWNPSVDLLASGS 188
                                     :::::|:::|:::|:::|
Db 1322 NGGTIGSASI-----DAILKWSPGLLILTKGHNSWNVSWVSFGPIRIFASGS 1371

```

189	QY	GDSTAR-WNLSENSTSGSTOLVLAHCHIRREGQDVPNSKOVTSITLDWNSGCTLLATGSDYDGF	248
1372	DB	RDXTVTLRWDE-----VLENPKGD-----NDWTSISFSGDETLLAASRDOT	1417
249	QY	AR1WTGDGNLASTLGQHGPIFALKNNKGNFILSAGVDKTTIINDAHTGEAKQOFPFHS	308
1418	DB	VKLSRHGKLNTFKGHTSGIWGVAWSPNQMTASAKOOTVKLWH-QDGKILHTLQGHQ	1476
309	QY	APALDWDVWOSNN-TFASCSTDMCHVCKLQGDPRPKTFQOHTNEVNAIKWDPTGNLLASC	367
1477	DB	DAYLAVAWSDSQVIAASAGKDKTIVKWSQG-GOLLHTLQOHTDAVKNWSFSPDGKLLASV	1535
368	QY	SDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPTGNPNANIMLASASDSTVRLW	427
1536	DB	SDDTTVKLMS-RDQQLLHTLKEHSRRVNGVAWSPDQG-----ILASASIDGTVKLW	1585
428	QY	DVDRGFCITLTKHQBPVTVSAFSPDGRVYLASGSPDKCVHIWN	470
1586	DB	NRD-GSLSRNLPGDGDSFTSVSFSFGKMLAANSDDO-IRLWN	1626

DEC 11 1966

RESOL1 8
 Ti18521
 beta transducin-like protein - Podospora anserina
 C:Species: Podospora anserina
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18521
 R:Saube, S.; Turcqi, B.; Begueret, J.
 Gene 162, 135-139, 1995
 A:Title: A gene responsible for vegetative incompatibility in the fungus Podospora anserina
 A:Reference number: Z18944; MUID:96009891; PMID:7557402
 A:Accession: T18521
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1356 <SAU>

A; Cross-references: EMBL: L28125; NID: g607002; PID: g607003; PIDN: AAA85775.1
C; Genetics:
A; Gene: Met-el
A; Introns: 761/3

Query Match 15.5%; Score 420.5; DB 2; Length 1356;
Best Local Similarity 28.2%; Pred. No. 5.2e-22;
Matches 117; Conservative 63; Mismatches 164; Indels 71;

[illegible]

RESULT 9

AD1842
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AD1842
R:Kansoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1711 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB77808.1; PID:g17135262; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: al10284

Query Match
15.0%; Score 408; DB 2; Length 1711;

QY	65	SINSDGTL----	FDGRIEISL----	LDIVMPDVGWTRQCAVYDKLACQAAAAAA	114
Db	1246	SASDDGIRLWSLDRGLIIP	PSHTQVLAVFSPD-----	GGTIVSA	1288
QY	115	AAAAQQGSAKGE-	NTANGENG-----	AHTIANHHTDMEVDGVEIPP	159
Db	1289	GADNTVKLWSRNGTLLTLE	GHNAAVWQVIFPDGRGLIATASADXTITW	SRDGN-----	1344
QY	160	NKAVVLRGHSEVFICAWN	PVSDLLASGSGDSTARIW-----		196

```
Db 1345 --LGTAGHNEVNSLSFSPDGNILASGSDNTVRLWNTLPTKTFYGHKGSVVRPS 1402
QY 197 NUGENTSGTQVLVLRHCIREG-----GQDVPSNKDVTSLDWSNSETLLATGSDVDFAR 250
Db 1403 NDKKITISLSTSTMTWSLDGKLLQTLASPLP--DVTISFTPDNKIVALASPDHTH 1459
QY 251 IWTKGNLASTLGHKGPIFALKWKKGNFILLSAGVDKTTIWDHTGAEAKQFPFHSAP 310
Db 1460 LYNROGLRLSPGHNHTISLSPNKQILASGSADKTIKWSVN-GRLLKTLIGHNGW 1518
QY 311 ALDVDQSN-NTFASCTDMCHVCKLQGDRIPIKTFQHTNEVNAIKWDPTGNLLASCSD 369
Db 1519 VTDIKESADGNIVASADKTIKWSL-DGRLIRTLQCHSASVWSVNLSPDQTLASTSQ 1577
QY 370 DMTLKWSMKQDNCVHDLQHNKEIYTIKWSPTGPTNNPNANILMASASPDSTVRLWDY 429
Db 1578 DETIKLWNNGE-LIYTLRHGSDVVNLSFSPDGK-----TIASASDGTIKLWNV 1627
QY 430 DRGICHTLTKHQEPYVSAFSDGRLASGDFDKCVHIWNTQ-VCLHYINGQVILN 485
Db 1628 PNGTLKTFQHRGGVRSVSPDGKILASGGHDTTVKWNLEGIELQTLNDELIN 1684

RESULT 10
AGI889
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AGI889
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; UID:21595285; PMID:11759840
A:Accession: AGI889
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-934 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072622.1; PID:g17130010; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0664

Query Match 14.9%; Score 404.5; DB 2; Length 934;
Best Local Similarity 24.1%; Pred. No. 4.4e-21;
Matches 124; Conservative 104; Mismatches 176; Indels 111; Gaps 19;

QY 8 VNFLVRYLOESGFSSAFTFGIKSHISQSNINGALVPPAALISIIQKLVYAEVGIN 67
Db 231 ISSLVTLILAGLAVLQGGKARMSETK-----AISSSAESFLNANLEF- 277
QY 68 EDGTLFD---GRPIESLSLIDA-VMPDVVQTRQAY-----RDKLAQQAQAAAAA 118
Db 278 -DGLIASIRARRIKGTGDIGANTRQITETIQQSINFVREKNRLAEHGMLESVSPD 336
QY 119 SQGSNKGENTANGENGAHTIANHNTDMEVDGVEIPPNKAVLVRGHESEVP-ICAW 177
Db 337 SXFIATASDKTVK-----IWSLDGKKQL-----VYLREEKEGENSVAF 376
QY 178 NPVSDLLASGSDSTARIWNLSNSTSGTQLVLRHCIREGQ--DVPSNKD-VTSLDWN 234
Db 377 SPDGTLMTGSDWNTAKIWS-----REGRLHTLDGKHAIVLEAFS 418
QY 235 SEGTLATGSDGFARITWKDGNLSTLQGHKGPFAKWNKGNFILLSAGVDKTTIWD 294
Db 419 PDSQLLATASWDNTVRLWSREGKLLHTLEGHKDKVNSITFSPDGLIATVGDNTWKLWN 478
QY 295 AHTGEAKQFPFHSAPALVDWQ-----SNNTFASCTDMCHVCKLQGDRIPIKTFQHT 349
Db 479 L-DGKELIRFRGHQ-----DMIVSVSFPDGKQIATASGDRTVKWSL-DGKELQIRGHQ 532
QY 350 NEVNAIKWDPTGNLLASCSDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGPG- 405
```

```
Db 533 NGVNSVTFSPDGKLIATASGDRITVKIWSKQGE-LETLYGHTDAVNSVAFSPDGTSTATA 591
QY 406 -----TNNPNANILM-----LASASFDSTVRLWDV--DRGICI 435
Db 592 GNDKTAKIWLKNSPNSIIVRGHEDEVDLVFSPNGKYIATASWKTAKLWSIVGDKLQEL 651
QY 436 HTLTKHQEPYVSAFSDGRLASGDFDKCVHIWN 470
Db 652 RTFNGHQVRNKLSPSPDGKYIATTSWDKTAKLWN 686

RESULT 11
AC1842
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC1842
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; UID:21595285; PMID:11759840
A:Accession: AC1842
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1747 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077807.1; PID:g17135261; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0283

Query Match 14.5%; Score 395; DB 2; Length 1747;
Best Local Similarity 28.2%; Pred. No. 5.1e-20;
Matches 105; Conservative 59; Mismatches 131; Indels 78; Gaps 9;

QY 165 LRHSEVFIICAMPVSDLLASGSDSTARINWL-----SENSTSGSTQ 208
Db 1310 LTGHERITSVKFSFDGKILASASGDKTIKFWNTDGKFLKLTIAAHNQVNSINFSPDSKT 1369
QY 209 LV-----LRHCIREGGQDV-----PSNK----- 226
Db 1370 LVSGAGSTWKMVKIDGTILKTISGRGEQRDVTFSPDNKVIASASDKTVIRQLNYQK 1429
QY 227 ----DVTSLDWSNSETLLATGSDGFARITWKDGNLA-----STLQGHKGPFAKWNKKG 278
Db 1430 SKSNVNSVSNPDGKTFASAGWDGNITWQRE-TLAHSSLSITIQKNQIITVSYSPDG 1488
QY 279 NFILSAGVDKTTIWDHTGAEAKQFPFHSAPALVDWQ--SNNTFASCTDMCHVCKLQ 337
Db 1489 XTIATASADNTIKLWDSQTQOLIKTLTGKDRITTLTSHFDPNQTIASGSADKTIKIRVN 1548
QY 338 QDRPIKTFQGHTEVNAIKWDPTGNLLASGSDMTLKIWSMKQDNCVHDLQHNKEIYTI 397
Db 1549 DGQLRLITGHNDVTSVNFSPDQFLASGSTNTVKI-W-QTDRLLIKNITGHGLATSV 1607
QY 398 KWSPTGCTNNPNANILMASASPDSTVRLWDVDRGICHTLTKHQEPYVSAFSDGRL 457
Db 1608 KFSPD-----SHTLASASWDNTIKLQWVTDGKLNNLNGHIDGVTSLSFSPDGEIL 1658
QY 458 ASGSDFDKCVHIWN 470
Db 1659 ASGSADNTIKLWN 1671

RESULT 12
S76086
beta transducin-like protein, 190K - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein sl0163
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 26-May-2000
```

C;Accession: S76086
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shingo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, S.
 DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S76086
 A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1693 <XAN>

A;Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PID:BA10064.1; PID:d101071
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Superfamily: unassigned WD repeat proteins; WD repeat homology

F;1051-1084/Domain: WD repeat homology <WD01>
 F;1092-1125/Domain: WD repeat homology <WD02>
 F;1133-1166/Domain: WD repeat homology <WD03>
 F;1174-1207/Domain: WD repeat homology <WD04>
 F;1256-1289/Domain: WD repeat homology <WD05>
 F;1297-1330/Domain: WD repeat homology <WD06>
 F;1338-1371/Domain: WD repeat homology <WD07>
 F;1420-1453/Domain: WD repeat homology <WD08>
 F;1461-1494/Domain: WD repeat homology <WD09>
 F;1502-1535/Domain: WD repeat homology <WD10>
 F;1584-1617/Domain: WD repeat homology <WD11>
 F;1625-1658/Domain: WD repeat homology <WD12>

Query Match 14.5%; Score 394.5; DB 2; Length 1693;

Best Local Similarity 28.2%; Pred. No. 5.3e-20;

Matches 115; Conservative 55; Mismatches 149; Indels 89; Gaps 12;

QY 100 RDLAQOQAAAAAASQGSAXNGENTANGE--ENGAHTIANNHTDMEVDGDI 157
 DB 1362 RDKTARLWTTGECVAVLADHQGVREGQSPDGQWIVTGS---ADKTAQLNVLG---- 1414

QY 158 PPKAVVLRGHSEVFICANPNVSDLLASGSGSTARIWNLSNSTGSLVLRHCIRE 217
 DB 1415 --KCLTVLRGHQDAVLNVRSPDSQIVTASDKGTARVN----NTGRELAVLRH---- 1463

QY 218 GGGVPSNKDVSLDWNSEGTLIATSGYGFARITWKGDLASTLQKHGPTFALKWKK 277
 DB 1464 -----YEKNIFAAEFASDQGFIVTASDDNTAGIWEIVREGVICRGHEGVPYFAQFSAD 1517

QY 278 GNFILSAGVDKTIINDAHTGEAKQFPFHSAPALVDQSNNTFASCTDMCIHVCKLG 337
 DB 1518 SRYILTSASVDNTARIWDF-----LTHNALVSP 1684

QY 338 QDRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIMSKMKNQDNCVHDLQHNKEIYTI 397
 DB 1538 --RPLLTLAGHQSIYVQARFSPENGLIATVSADHTARLWD-RSGKTAVLYGHQGLVGTV 1594

QY 398 KNSPTGCTNNPNANMLASASFDSTVRLWDVRGICHTLTKHQBPVYSVAFSPDGRYL 457
 DB 1595 DWSPDGO-----MLVTASNDGTARLWDLG-GRELLTLEBGHWNVSARESPDGRWV 1644

QY 458 ASGSFDCVCHWNTQVCLHYLN--GOVLLNLRGSIICLYTLPHLHVIP 503
 DB 1645 LTSSADGTAKLPVKTLPLQLSQGGWLKNY-----LTHNALVSP 1684

RESULT 13

T40883

WD repeat protein - fission yeast (*Schizosaccharomyces pombe*)

C;Species: *Schizosaccharomyces pombe*

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C;Accession: T40883

R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.

submitted to the EMBL Data Library, September 1998

A;Reference number: Z21954

A;Accession: T40883

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-564 <WOO>

A;Cross-references: EMBL:AL031764; PIDN:CAA21113.1; GSPDB:GN00066; SPDB:SPCC1235.09
 A;Experimental source: strain 972h-; cosmid c1235

C;Genetics:

A;Gene: SPDB:SPCC1235.09

A;Map position: 3

A;Introns: 18/1; 273/3; 413/3

Query Match 14.1%; Score 382.5; DB 2; Length 564;

Best Local Similarity 24.4%; Pred. No. 8.5e-20;

Matches 131; Conservative 106; Mismatches 198; Indels 101; Gaps 24;

QY 3 ISSDEVNPLVRYLOESGFSHSATFGIKSHISOSNIN---GALVPPAALISIIQKGLQY 59
 DB 1 MDTQVNYIIRYLYKEGYSHTKFAFERETGI---QNLDKQWGSTCQGVALVELQKGLQY 58

QY 60 VEAEV-----SINEDG--TLFGDRPIESLSLD-----AVMPDVVQT----- 94
 DB 59 VELEKHYVDNHSSNEBASKTSDGE-----SLVNNENCKLPFLYLVPHICETTLTKADST 113

QY 95 -----RQAYRKLAQOQAAAAAASQGSAXNG-----ENTANGENGHTIA 142
 DB 114 NGFCEHNSNDHQLKILQDKGSGSPSPVMPFDDKIEKRDIIDITMADESINVKDAPRIA 173

QY 143 NNHTDMEVDGDVEIPNKAVVLRGHE---SEVF-----ICA-WNPVSD-----LLA 185
 DB 174 VYNSPV-----TETIKQVFTTGGEDIKSDFFKVIPTKHPVTCADWRPQLQENHYVE 228

QY 186 SSGSGSTARIWNLS-----ENSTGSLVLRHCIREGGQVPSNKDVSLDWNSEGTLLA 241
 DB 229 FSIQMTNATLASVTCBEQNDFKAKTD-----YCL-----QSSFDNQDITGVANNNSGFLA 280

QY 242 TGSVDGPFARIWTKDGNLSTLQGHKGFIFALKNNKKNFILSAGVDKTIINDAHTGEAK 301
 DB 281 YAFSGVIEIYDSHGSQLSPHNKKGVLKXSGTITYLAAGSADGTITLFD-----QLK 336

QY 302 Q-QPFPH--SAPALVDVQSNNTFASCTDMCIHVCKLQGDRIKTF-QGHTEVNAIKW 357
 DB 337 QTYQSIDTLASSVLDIEWISFDEFVTSDEGLSVYKVDGKAPVSTVSHADNSIVALRY 396

QY 358 DPTGNLLASCSDDMTLKIMSKQD---NCVHDLQHNKEIYTIKWSPTGPTNNPNANLM 414
 DB 397 NLRSLLLTASDSTVTKLWSRGDAGAECLH-VPSFSPVNCIDW-----NURGTPI 448

QY 415 LASASFSTVRLWDVRGICHTLTKHQBPVYSVAFSPDGRYLASGSFDCVCHW 470
 DB 449 LAVAS-NSIVGMYNAISLQQLAVPMRHTAPVSALSFHNGRYLATGDTSGGVCWS 503

RESULT 14

AB2410

WD-repeat protein [imported] - *Nostoc* sp. (strain PCC 7120)

C;Species: *Nostoc* sp. PCC 7120

A;Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AB2410

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AB2410

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1551 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA076533.1; PID:g17133971; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: all4834

Query Match 14.0%; Score 381; DB 2; Length 1551;

Best Local Similarity 23.4%; Pred. No. 4.4e-19;

Matches 118; Conservative 94; Mismatches 175; Indels 118; Gaps 18;

```
QY 48 ALISIIKGLQYVEAEVINEGDTLFDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQQ 107
Db 903 ALMSAVRSQ-----KALQALVKGRSLAKYPAISPLALOTILDNIOERNQ-----FQGHQ 953
QY 108 AAAAAAASQGSAXGENTANGENGHAHTIANHHTDMMEVDGVEIPPNAKAVILRG 167
Db 954 AWRSVSFSRGGQYIILASDDCTAR-----LWNLQ-----KQLISLQG 992
QY 168 HESEVFICAWNPVSDLLASGSDSTARIWNLSNSTS-----GSTQLVL-----RHCIR 216
Db 993 HEDTIWSANFSPGKVIATASSDRTARLWNFGQQALAKFQGHQYVRSVSPDGKHIAT 1052
QY 217 EG-----GQDV-----PSNKDV-TSLD-----WNSEGTL-- 240
Db 1053 AGDHTARLWFSFGQQLVOPFGHQGTVMCISFSPDGKHIATAADDRIVELWNLKGLLVR 1112
QY 241 -----ATGSGYGFARIWTKGNLASTLGGHKGPIFALKWKNKGNF 280
Db 1113 FPGHQCVMDVSPSPSQVIATASSDGTSLRLNLAGEQITRFGHGGVMSVRSFSPNGQY 1172
QY 281 ILSAGVDKTTIIWDAHTGEAKQFPFHSAPALDWDVQSNNTF-ASGSTDVDCIHVCKLGOD 339
Db 1173 IATSSDRTARVWNLN-GQLAQFSGHQDYVRSVSPDGKVIATASSDRTVRLWHLNKG 1231
QY 340 RPKTFQGHTEVNAIKWDPGTGNLLASCSDMTLKIWSMKQDNCVHDLQQHKEIYTIKW 399
Db 1232 Q-FSAFQGHQSTVRSVDFSPDGQKVTAADDRTVRLWNLKGEELLQFL-GHRGKVMVSF 1289
QY 400 SPTGPGTNNPNANMLASAFSTVRLWDVDRGICHTLTKEHQPVSVAFSFSPDGRYLAS 459
Db 1290 SDGK-----YIATSSDRTVRLWDI-TGOLLQPPGHQGTVMVSFSPDGQHIAT 1339
QY 460 GSPDKCVHIWNTQVCLHYLNGQVLL 484
Db 1340 ASSDLTTRLWS-----LDGQELM 1357

RESULT 15
AH2195
hypothetical protein alr3119 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH2195
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11753840
A:Accession: AH2195
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-676 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074818.1; PID:gl7132214; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3119

Query Match 13.6%; Score 369; DB 2; Length 676;
Best Local Similarity 23.7%; Pred. No. 1e-18;
Matches 116; Conservative 96; Mismatches 194; Indels 84; Gaps 13;

QY 21 FSHSAFTFGIKSHISOSNINGALVPAPALISIIQGLQYVEA-----EVS 65
Db 226 FSLGATCFHLLTGINPSNL-----FVEQYSWVESWQQYWNFTNSDRNEGYL 273
QY 66 INEDGTLFDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQQAAAAA-----AAAAASQGG 122
Db 274 VKVLNKLLETDIQRYYQSADAEVWMDLTKQSLSLRKLTTIPKSAIFSRWSASTSLTAST 333
QY 123 SAKNGENTANGENGHAHTIANHHTDMMEVDG-----DVEIPPNAKAVILRG 167
Db 334 TTKQAWKLNGRLK-QOLLINTMSALLGLVGVGHLSQPLITKFSBISTQP---YTLKG 389
```

```
QY 168 HESEVFICAWNPVSDLLASGSDSTARIWNLSNSTSGSTOLVLRHCIREGGQDVPSNKG 227
Db 390 HASDVNSVAFSPNGEFLASGSDDKTIKVMNLKN-----KQKIHTLPQHS 434
QY 228 -VTSLDWNSEGTLATGSDGFARIWTKDGNLAS-----TLGQHKGPFIKWKNNKGNFI 281
Db 435 WWAIAFSPDGKTLASTGADXTIKLW-----NLATGKEIRHLKGHSQGVASVAFSPDGKTL 490
QY 282 LSAGVDKTTIIWDAHTGEAKQFPFHSAPALDWDVQ-SNNTFASCSCTDMCIHVCKLGQDR 340
Db 491 ASGLDKTIKLNWPAATKEIRTLQEHSSGVANVAFSPDGKTLASGSDKTIKLNLTTSK 550
QY 341 PIKTFQGHTEVNAIKWDPGTGNLLASCSDMTLKIWSMKQDNCVHDLQQHKEIYTIKW 400
Db 551 VIHTLKHSDDLUMSVAFNSDQTLASGSKTKIKLWNLSTGKTIRTLRGHSDKVNVAIV 610
QY 401 PTGPGTNNPNANMLASAFSTVRLWDVDRGICHTLTKEHQPVSVAFSFSPDGRYLAS- 459
Db 611 -----PRDSTVLAGSNDNTIKLWNLTTGEIRTLKRDSGYIYSIVISPDGRNLASG 662
QY 460 GSPDKCVHIW 469
Db 663 GSAENIIRKIW 672
```

Search completed: August 9, 2004, 16:48:49
Job time : 14.3333 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:40:42 ; Search time 8.66667 Seconds
(without alignments)
3088.161 Million cell updates/sec

Title: US-09-987-701-4
Perfect score: 2718
Sequence: 1 MSISDEVNFLVRYLQESG.....LPHHLWTPLVALTELLVVK 514

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2493	91.7	514	1	TBLR1_HUMAN
2	2229	82.0	526	1	TBLX_HUMAN
3	2170	79.8	522	1	TBLX_HUMAN
4	1308	48.1	313	1	TBLX_MOUSE
5	475	17.5	1526	1	YV46_ANASP
6	472	17.4	1335	1	YF22_YEAST
7	458.5	16.9	1258	1	YS00_ANASP
8	451.5	16.6	1683	1	YL24_ANASP
9	420.5	15.5	1356	1	HET1_PODAN
10	394.5	14.5	1693	1	Y163_SYNY3
11	366.5	13.5	485	1	HU87_HUMAN
12	365.5	13.4	485	1	HU87_MOUSE
13	351.5	12.9	409	1	LI81_HUMAN
14	350.5	12.9	409	1	LI81_BOVIN
15	350	12.9	339	1	CI41_HUMAN
16	349.5	12.9	409	1	LI81_MOUSE
17	339	12.5	800	1	T2D4_HUMAN
18	337	12.4	515	1	YCW2_YEAST
19	333	12.3	742	1	PKA3_THECU
20	331.5	12.2	361	1	WD3_DROME
21	331	12.2	437	1	AAC3_DICDI
22	325.5	12.0	334	1	WD85_HUMAN
23	319.5	11.8	606	1	PF20_CHLRE
24	317	11.7	579	1	SE10_CAEEL
25	310	11.4	605	1	POF1_SCHPO
26	309.5	11.4	376	1	YKX4_CAEEL
27	308	11.3	704	1	T2D4_YEAST
28	307	11.3	798	1	LI23_CAEEL
29	305	11.2	665	1	YH92_CAEEL
30	300	11.0	501	1	YH92_CAEEL
31	299	11.0	640	1	MT30_YEAST
32	291	10.7	1191	1	Y143_SYNY3
33	289.5	10.7	395	1	GBB5_HUMAN

34	288.5	10.6	586	1	TU12_SCHPO
35	287.5	10.6	395	1	GBB5_MOUSE
36	286	10.5	678	1	SCOB_EMBNI
37	285	10.5	605	1	FW1A_HUMAN
38	283	10.4	395	1	YZLL_CAEEL
39	283	10.4	542	1	FW1B_HUMAN
40	283	10.4	931	1	LUG_ARATH
41	282.5	10.4	682	1	TU11_KLULA
42	281	10.3	554	1	PRP4_ARATH
43	280.5	10.3	380	1	GBB_MAIZE
44	280	10.3	614	1	TU11_SCHPO
45	279	10.3	522	1	PRP4_HUMAN

ALIGNMENTS

RESULT 1					
TBLR1_HUMAN					
ID	TBLR1_HUMAN	STANDARD;	PRT;	514	AA.
AC	Q9BZK7; Q9H211; Q9H9A1;				
DT	28-FEB-2003 (Rel. 41, Created)				
DI	28-FEB-2003 (Rel. 41, Last sequence update)				
DE	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Nuclear receptor co-repressor/HDAC3 complex subunit TBLR1 (TBL1-related protein 1).				
GN	TBLR1 OR IRA1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND INTERACTION WITH HDAC3.				
RX	MEDLINE=21930350; PubMed=11931768;				
RA	Zhang J., Kalkum M., Chait B.T., Rosder R.G.;				
RT	"The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2.";				
EL	Mol. Cell 9:611-623(2002).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Zhang X., Dormady S., Basch R.;				
RT	"Identification of four human cDNAs that are differentially expressed by early hematopoietic progenitors.";				
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Aotaka S., Yoshikawa Y., Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y., Sasaki N.;				
RT	"NED0 human cDNA sequencing project.";				
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.				
CC	-1- SUBUNIT: Interacts with HDAC3.				
CC	-1- SIMILARITY: Contains 1 Lish domain.				
CC	-1- SIMILARITY: Contains 8 WD repeats.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
DR	EMBL; AF314544; ARK00301.1; -				
DR	EMBL; AF268193; AAG44736.1; -				
DR	EMBL; AK022956; BAB1431.1; -				
DR	InterPro; IPR006594; Lish.				
DR	InterPro; IPR01680; WD40.				
DR	Pfam; PF00400; WD40; 8.				
DR	PRINTS; PR00320; GPROTEINRPT.				
DR	ProDom; PD000018; WD40; 3.				
DR	SMART; SM00657; Lish; 1.				

DR SWART; SM00320; WD40; 8.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00082; WD_REPEATS_2; 6.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT DOMAIN 4 36 LISH.
 FT REPEAT 167 206 WD 1.
 FT REPEAT 223 262 WD 2.
 FT REPEAT 264 303 WD 3.
 FT REPEAT 306 344 WD 4.
 FT REPEAT 347 386 WD 5.
 FT REPEAT 389 437 WD 6.
 FT REPEAT 440 479 WD 7.
 FT REPEAT 481 513 WD 8.
 FT DOMAIN 108 118 POLY-ALA.
 FT CONFLICT 31 31 E -> K (IN REF. 2).
 FT CONFLICT 59 59 Y -> H (IN REF. 3).
 FT CONFLICT 389 389 A -> Q (IN REF. 2).
 SQ SEQUENCE 514 AA; 55595 MW; 0855622EE4BA796D CRC64;

Query Match 91.7%; Score 2493; DB 1; Length 514;
 Best Local Similarity 99.6%; Pred. No. 2.4e-174;
 Matches 470; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSISSEVFLVRYLQSGFHSFTFGIKHSISOSNINGALVPPAALISIIQKGLQV 60
 DB 1 MSISSEVFLVRYLQSGFHSFTFGIESHSISOSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSNEDGTLFDGRIEISLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 DB 61 EAEVSNEDGTLFDGRIEISLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 QY 121 QGSAKNGENTANGAHTIANNHDDMEVDGVEIPENKAVLGRGHESEVFIKANPV 180
 DB 121 QGSAKNGENTANGAHTIANNHDDMEVDGVEIPENKAVLGRGHESEVFIKANPV 180
 QY 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGQDVPSKDVTSLDWNSGTL 240
 DB 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGQDVPSKDVTSLDWNSGTL 240
 QY 241 ATGSDGFARITWTDGNLSTLQHGKPIFALKWKKGNFILLSAGVDKTIWDHATGEA 300
 DB 241 ATGSDGFARITWTDGNLSTLQHGKPIFALKWKKGNFILLSAGVDKTIWDHATGEA 300
 QY 301 KQFPFHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
 DB 301 KQFPFHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKWSMKQDNCVHDLQHNKEIYTIKSPGPGTNNPNNMLLASASF 420
 DB 361 GNLLASCSDDMTLKWSMKQDNCVHDLQHNKEIYTIKSPGPGTNNPNNMLLASASF 420
 QY 421 DSTVRLWDVDRGCIHTLTKHQBPVSVAFSPDGRVYLASGDFKCVHIWNTQ 472
 DB 421 DSTVRLWDVDRGCIHTLTKHQBPVSVAFSPDGRVYLASGDFKCVHIWNTQ 472

RESULT 2
 TBXLX HUMAN STANDARD; PRT; 526 AA.
 AC 060907;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transducin beta-like 1x protein (Transducin-beta-like 1, X-linked).
 GN TBXLX OR TB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=99264241; PubMed=10330347;
 RA Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,
 RA Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.,
 RT "X-linked late-onset sensorineural deafness caused by a deletion
 RT involving OAI and a novel gene containing WD-40 repeats.";
 RL Am. J. Hum. Genet. 64:1604-1616(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph.
 RX MEDLINE=2386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Rodriguez A.C., Touchman J.W., Green E.D., Dickson M.C.,
 RA Blakesley R.W., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schemm A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- SIMILARITY: Contains 1 Lish domain.
 CC -!- SIMILARITY: Contains 8 WD repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch.
 CC EMBL; Y12781; CAA73319.1; ALT_INIT.
 DR EMBL; EC032708; AAB32708.1; -.
 DR Genew; HGNC:11585; TBLIX.
 DR MIM; 300196; -.
 DR GO; GO:0007605; P:hearing; TAS.
 DR GO; GO:0007165; P:signal transduction; TAS.
 DR GO; GO:0007601; P:vision; TAS.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 7.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00082; WD_REPEATS_2; 6.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT DOMAIN 4 36 LISH.
 FT REPEAT 179 218 WD 1.
 FT REPEAT 235 274 WD 2.
 FT REPEAT 276 315 WD 3.
 FT REPEAT 318 358 WD 4.
 FT REPEAT 359 398 WD 5.
 FT REPEAT 401 449 WD 6.
 FT REPEAT 452 491 WD 7.
 FT REPEAT 493 525 WD 8.
 FT DOMAIN 108 124 POLY-ALA.
 SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F6E9 CRC64;


```
Query Match      82.0%; Score 2229; DB 1; Length 526;
Best Local Similarity 85.4%; Pred. No. 4.4e-155;
Matches 414; Conservative 32; Mismatches 25; Indels 14; Gaps 2;

QY 1 MSISDEVNLFVLYLQESGFSHSATFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
DB 1 MSISDEVNLFVLYLQESGFSHSATFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
QY 61 EAEVINEGTLDPGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQOQAAAAA----- 117
DB 61 EAEVINEGTLDPGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQOQAAAAA----- 117
QY 118 -----ASOQGSAGKNGENTANGENGATIANHNTDMVEVDGVEIPNKAIVLRG 167
DB 121 ATAAITTSAGVSHQNSKREATVNGEENRAHSV-NNHAKPMEIDGVEIPSSKATVLRG 179
QY 168 HSEVFICAWNPNVSDLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPNSKD 227
DB 180 HSEVFICAWNPNVSDLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPNSKD 239
QY 228 VTSLDVNSGTLTATGSGYGFARIWTKDGNLSTLQGHKGPFPALKWNKKNFILLSAGVD 287
DB 240 VTSLDVNTGTLTATGSGYGFARIWTKDGNLSTLQGHKGPFPALKWNKKNFILLSAGVD 299
QY 288 KTTIWDATHTGEAKQOQFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQG 347
DB 300 KTTIWDATHTGEAKQOQFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQG 359
QY 348 HINEVNAIKWDPTGNLLASCDDMTLKIWSMKODNCVHDLQHNKEIYTIKWSPTGPGNN 407
DB 360 HINEVNAIKWDPTGNLLASCDDMTLKIWSMKQEVCIHDLQHNKEIYTIKWSPTGPGNS 419
QY 408 NPNANMLASAFSDSVRLVDVDRG:CIHTLTKHQEPVSVAFSPDGRVLAGSGSPDKCVH 467
DB 420 NPNANMLASAFSDSVRLVDVDRG:CIHTLTKHQEPVSVAFSPDGRVLAGSGSPDKCVH 479
QY 468 INWNTQ 472
DB 480 INWNTQ 484

RESULT 3
TEBU_HUMAN
ID TEBU_HUMAN STANDARD; PRT; 522 AA.
AC Q9BQ87;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE "The DNA sequence of the human Y chromosome."
DE Transducin beta-like 1 protein (transducin-beta-like 1, Y-linked).
GN TEBU OR TEBU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown J.G., Rozen S.,
RA Wilson R.K., Waterston R.H., Page D.C.;
RT "The DNA sequence of the human Y chromosome."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF332220; AAK13472.1; -.
DR EMBL; AF332221; AAK13473.1; -.

```

```
DR EMBL; AF332222; AAK13474.1; -.
DR Genew; HGNC:18502; TBLIY.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PRO0320; GPROTEINBRPT.
DR PRODOM; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS00294; WD REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT DOMAIN 4 36 LISH.
FT REPEAT 177 216 WD 1.
FT REPEAT 233 272 WD 2.
FT REPEAT 274 313 WD 3.
FT REPEAT 316 354 WD 4.
FT REPEAT 357 396 WD 5.
FT REPEAT 399 447 WD 6.
FT REPEAT 450 489 WD 7.
FT REPEAT 491 521 WD 8.
SQ SEQUENCE 522 AA; 56688 MW; 4E020216422442D8 CRC64;

Query Match      79.8%; Score 2170; DB 1; Length 522;
Best Local Similarity 84.1%; Pred. No. 8.7e-151;
Matches 406; Conservative 32; Mismatches 33; Indels 12; Gaps 2;

QY 1 MSISDEVNLFVLYLQESGFSHSATFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
DB 1 MSISDEVNLFVLYLQESGFSHSATFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
QY 61 EAEVINEGTLDPGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQOQAAAAA----- 115
DB 61 EAEVINEGTLDPGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQOQAAAAA----- 115
QY 116 -----AAASQGSAGKNGENTANGENGATIANHNTDMVEVDGVEIPNKAIVLRGHE 169
DB 121 AATWTPAAISQNPFPKREATVNGEENGATIANHNTDMVEVDGVEIPNKAIVLRGHE 179
QY 170 SEVFICAWNPNVSDLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPNSKDV 229
DB 180 SEVFICAWNPNVSDLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPNSKDV 239
QY 230 SLDWSNSETLLATGSGYGFARIWTKDGNLSTLQGHKGPFPALKWNKKNFILLSAGVDKT 289
DB 240 SLDWSNSETLLATGSGYGFARIWTKDGNLSTLQGHKGPFPALKWNKKNFILLSAGVDKT 299
QY 290 TTIWDAHTGEAKQOQFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGH 349
DB 300 TTIWDAHTGEAKQOQFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGH 359
QY 350 NEVNAIKWDPTGNLLASCDDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGPGNNP 409
DB 360 NEVNAIKWDPTGNLLASCDDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGPGNNP 419
QY 410 NANMLASAFSDSVRLVDVDRG:CIHTLTKHQEPVSVAFSPDGRVLAGSGSPDKCVHIW 469
DB 420 NSSIMLASAFSDSVRLVDVDRG:CIHTLTKHQEPVSVAFSPDGRVLAGSGSPDKCVHIW 479
QY 470 NTQ 472
DB 480 NTQ 482

RESULT 4
TEBLX_MOUSE
ID TBLX_MOUSE STANDARD; PRT; 313 AA.
AC Q9QX67;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

Transducin beta-like IX protein (Fragment).
DE TBL1X OR TBL1.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=129/SVJ;
RC Botcherby M.R.M., Straw R., Clarke D., Greystrom J.S., Weston P.,
RA Hunter G., Kimberley C., Rhodes M.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains at least 5 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; F38006; CAB61534.1; -
CC MGD; MGI:1336172; Tbl1x.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 5.
CC PRINTS; PR00320; GPROTEINRPT.
CC ProDom; PD000018; WD40; 2.
CC SMART; SM00320; WD40; 5.
CC PROSITE; PS00678; WD REPEATS 1; 3.
CC PROSITE; PS0082; WD REPEATS 2; 4.
CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT NON TER 1 1 WD 1.
FT REPEAT <1 24 WD 2.
FT REPEAT 76 115 WD 3.
FT REPEAT 160 198 WD 4.
FT REPEAT 201 249 WD 4.
FT REPEAT 252 291 WD 5.
FT NON TER 313 313
SQ SEQUENCE 313 AA; 34808 MW; 3BC48683432DFB7 CRC64;
Query Match 48.1%; Score 1308; DB 1; Length 313;
Best Local Similarity 83.2%; Pred. No. 3.2e-88;
Matches 243; Conservative 16; Mismatches 19; Indels 14; Gaps 4;
Qy 188 SGTSTARIWNLSENSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSG-----LLAT 242
Db 2 SGTSTARIWNLSENSTSGTQVLVLRHCIREGGHDFVSNKDVTSLDWNEHCWQVPMVL 61
Qy 243 GSYDGFARITWKGNLASTLQGHKGPFPFALKWKNKGNFILSAGVDKTTIWDHTGEAKQ 302
Db 62 QEYQ-----QKSNLASTLQGHKGPFPFALKWKNKGNFILSAGVDKTTIWDHTGEAKQ 114
Qy 303 QFPFHS-APALVDVQSNNTFASCTDMCHVCKLGDRIKTFQGHTEVNAIKWDPTG 361
Db 115 QFPFHSAPALVDVQSNNTFASCTDMCHVCKLGDRIKTFQGHTEVNAIKWDPTG 173
Qy 362 NLLASCSDDMTLTKWSKQNCVHDLQHNKEIYTIKWSPTGPTNPNANMLASASFD 421
Db 174 MLLASCSDDMTLTKWSKQNCVHDLQHNKEIYTIKWSPTGPTNPNANMLASASFD 233
Qy 422 STVRLWDVDRGICHTLTKQHPYVSAFSDGRLASGFDKCVHWTQV 473
Db 234 STVRLWDVDRGICHTLTKQHPYVSAFSDGRLASGFDKCVHWTQV 285
RESULT 5
Y46 ANASP STANDARD; PRT; 1526 AA.
ID Y46 ANASP
AC Q8YR11;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein alr3466.
GN ALR3466.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: Contains 16 WD repeats.
CC -!- SIMILARITY: Contains 1 pentapeptide repeat domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003593; BAB75165.1; -
CC FIR; AC2239; AC2239.
CC InterPro; IPR007111; NACHT NTPase.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 15.
CC PRINTS; PR00320; GPROTEINRPT.
CC ProDom; PD000018; WD40; 8.
CC SMART; SM00320; WD40; 15.
CC PROSITE; PS00678; WD_REPEATS 1; 11.
CC PROSITE; PS0082; WD_REPEATS 2; 15.
CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
CC PROSITE; PS0837; NACHT; UNKNOWN 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 334 376 WD 1.
FT REPEAT 823 862 WD 2.
FT DOMAIN 823 862 PENTAPEPTIDE.
FT REPEAT 862 901 WD 2.
FT REPEAT 904 945 WD 3.
FT REPEAT 946 985 WD 4.
FT REPEAT 988 1027 WD 5.
FT REPEAT 1030 1069 WD 6.
FT REPEAT 1072 1111 WD 7.
FT REPEAT 1114 1153 WD 8.
FT REPEAT 1156 1195 WD 9.
FT REPEAT 1198 1237 WD 10.
FT REPEAT 1240 1279 WD 11.
FT REPEAT 1282 1321 WD 12.
FT REPEAT 1324 1363 WD 13.
FT REPEAT 1366 1405 WD 14.
FT REPEAT 1408 1447 WD 15.
FT REPEAT 1450 1491 WD 16.
SQ SEQUENCE 1526 AA; 170376 MW; D5DF272509B4738 CRC64;
Query Match 17.5%; Score 475; DB 1; Length 1526;
Best Local Similarity 31.0%; Pred. No. 1.5e-26;
Matches 119; Conservative 63; Mismatches 138; Indels 64; Gaps 10;
Qy 136 NCAHTIANHTDMVEVDGVEIPNKAV-VLRGHESEVFICAWNPVSDLLASGSDSTAR 194
Db 1127 NGV-TLANGSSDQIVRLWD--ISSKKCLYTLQGHNTNNAVAFSPDGTATLASGSDQTVR 1183
Qy 195 IWLNSSENS-----TSGSTQLVLRHCIREGGQDVPSNKNVTSLDWNSGFTLLATGSDYDFAR 250
Db 1184 LWDISSSKCLYTLQHTSW-----VNSVVFNPDGSTLASGSDQTVR 1225
Qy 251 IW-TKGNLASTLQGHKGPFPFALKWKNKGNFILSAGVDKTTIWDHTGEAKQFPFHS 309

```

Dd 1226 LWEINSSKCLCTFOGHTSWNSVFNPDGSLAGSGDKVRLWLDISSSKLHTFOGHT- 1284
Qy 310 PALVDVQSNNTF-----ASCSTDMCIHVCKLQDRIKTFQGHTEYNAIKWDPGTGN 362
Dd 1285 -----NWVNSVAENPDGSLAGSGDQTVRLWEISSKCLHTFOGHTSWNSVTFSPDGT 1339
Qy 363 LLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKSPGPGTNNPNANMLASAFDS 422
Dd 1340 MLASGSDQDQVRLWISISGECLEFLHTNWNVGSVIFSPDG-----AILASGSDQ 1390
Qy 423 TVRLWVDRCIGIHTLTKHOEYVYVAFSPDGYLAGSPDKCVHWNTOVCLHYLNGQV 482
Dd 1391 TVRLWSTSGKCLYTLQGHNNWGSIVFSPDGTLLASGSDQDQVRLW----- 1437
Qy 483 LLNLGRSICLYTLPHLWVPLVA 506
Dd 1438 --NISGECLYTLGHINSVRSVA 1459

RESULT 6
SIF2 YEAST
ID SIF2 YEAST STANDARD; PRT; 535 AA.
AC P38262;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE SIF2-interacting protein SIF2.
GN SIF2 OR EMB1 OR YBR103W OR YBR0832.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9288C;
RX MEDLINE=95208357; PubMed=7900425;
RA Mannhaup G., Stucka R., Ehme S., Vetter I., Feldmann H.;
RT "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL Yeast 10:1363-1381(1994).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=98315485; PubMed=9651685;
RA Cockell M., Renauld H., Watt P., Gasser S.M.;
RT "Sif2p interacts with Sir4 amino-terminal domain and antagonizes
telomeric silencing in yeast.";
RL Curr. Biol. 8:787-790(1998).
RN [3]
RP IDENTIFICATION IN A COMPLEX WITH HOS2; HST1; SNT1; CPRI; YIL112W AND
SET3.
RX MEDLINE=21567937; PubMed=11711434;
RA Fijnappel W.W.M.P., Schaft D., Roguev A., Shevchenko A., Tekotte H.,
RA Wilm M., Rigaut G., Seraphin B., Aasland R., Stewart A.F.;
RT "The S. cerevisiae SET3 complex includes two histone deacetylases,
Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation
gene program.";
RL Genes Dev. 15:2991-3004(2001).
CC -!- FUNCTION: Antagonizes telomeric silencing in yeast. May recruit
CC SIR4 to non-telomeric sites or repression.
CC -!- SUBUNIT: Interacts with SIR4 amino-terminal domain. Interacts with
CC a complex composed of SIN3 and RPD3. Identified in the Set3C
CC complex with HOS2, HST1, SNT1, CPRI, HOS4/YIL112W and SET3.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC EMBL; X78993; CAA55606.1; --
DR EMBL; Z35972; CAA85058.1; --
DR PIR; S48268; S48268.
DR GerOnline; 138646; --
DR SGD; S000307; SIF2.
DR GO; GO:0000118; C:histone deacetylase complex; IPI.
DR GO; GO:0005634; C:nucleus; IPI.
DR GO; GO:0017136; F:NAD-dependent histone deacetylase activity; IDA.
DR GO; GO:0045129; F:NAD-independent histone deacetylase activity; IDA.
DR GO; GO:0006348; P:chromatin silencing at telomere; IGI.
DR GO; GO:0016575; P:histone deacetylation; IDA.
DR GO; GO:0045833; P:negative regulation of meiosis; IPI.
DR GO; GO:0000122; P:negative regulation of transcription from P...; IPI.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 5.
DR ProDom; PD000018; WD40; 1.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD REPEATS 1; FALSE_NEG.
DR PROSITE; PS0082; WD REPEATS 2; 4.
DR PROSITE; PS0294; WD REPEATS_REGION; 1.
KW Repeat; WD repeat; Chromatin_regulator; Nuclear protein.
FT DOMAIN 4 36 LISH.
FT REPEAT 163 208 WD 1.
FT REPEAT 218 257 WD 2.
FT REPEAT 259 298 WD 3.
FT REPEAT 315 354 WD 4.
FT REPEAT 357 396 WD 5.
FT REPEAT 399 437 WD 6.
FT REPEAT 440 479 WD 7.
FT CONFLICT 396 396 S -> C (IN REF. 1; CAA85058).
SQ SEQUENCE 535 AA; 59145 MW; 25DD19AB2BF84B07 CRC64;

Query Match 17.4%; Score 472; DB 1; Length 535;
Best Local Similarity 25.8%; Pred. No. 6.2e-27;
Matches 140; Conservative 91; Mismatches 230; Indels 82; Gaps 15;

Qy 1 MSISDEVNLFVRYLQSGFSHSA-----FTFGIKSHISOSNINGALVPPAAL 49
Dd 1 MSITSEELNYLMRYCQEMGHEVSALALQDTRVLEDFEYKEHI-----PLGTL 50
Qy 50 ISITQKQGVVEAEVSNINEDG---TLFDGRPIESLSLIDAWPDVQVTRQRAYDKLAQQ 106
Dd 51 VNLVQRGILYTESLWVSKGDISALNEHLSDFNLVQALQID-----KEKPEI 101
Qy 107 QAAAAAASOQGSAGKNGENTANGEANGAHTIANNHTDMMEVDGDEIIPPKNKAVYLR 166
Dd 102 SSEGRFTLETNSKAGEDCASTVERETQDDTNSIDSSD--DLDFGVKI--LKEIV-- 155
Qy 167 GHESEVFCANVPYSD--LLASGSGDSTARINWLNSENSTSGSQ-----LVLRHCIREGG 219
Dd 156 -KLDNIVSSWNPLDESILAYGEKNVARLARIVETDQEGKKYWKLTIIAELRPFALSA 214
Qy 220 QDVPSNKDVTSLDWNSECTLLATGSYDGFARIMTKQGNLASTLQGHKGPFAKWNKGN 279
Dd 215 SSGKTTNQVTCIAWSHDGNSIVTGVENGELRNKTKGALLNVLNFRAPIVSVKWNKDG 274
Qy 280 FILSAGYDKTIIIDWATGAKQFPF-----HSAP---ALDVMQSNNTFAS 324
Dd 275 HIISMDYENVITLWNVISGTVMQHFKETGGSSINAEHSGDGLGVDEWVDDDFVI 334
Qy 325 CSTDMCIHVCKLQDRIKTFQGHTEYNAIKWDPGTNNPNANMLASGSDQDQVRLW 384
Dd 335 PGPKGALFVQITEKTFGKLIGHGPISVLEFNDTNKLLSASDDGTLIHWGGNGNSQ 394
Qy 385 HDLQOHNKEIYTIKSPGPGTNNPNANMLASGSDQVRLWVDRCIGIHTLTKHOE 444
Dd 395 NSFYGHSSQSVASW-----GDDKVISCSMDGSLVRLSLKQNTLLALSIVDGV 444
Qy 445 VYVAFSPDGRYLASGSDKCVHWNTOVCLHYLNGQVRLW-RSICLYTLPHLWV 503
```

Db 445 IFAGRISQDGQKVAFAFMDGQVNVYD-----LKLNSKRSRSLYGRDGLNPLP-----IP 495

QY 504 LYA 506
Db 496 LYA 498

RESULT 7

YS00 ANASP
ID YS00 ANASP STANDARD; PRT; 1258 AA.
AC Q8YTC2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 29-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical WD-repeat protein alr2800.
GN ALR2800.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.,
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- SIMILARITY: Contains 15 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003590; BAB74493.1;
CC PIR; A12155; A12155.
CC InterPro; IPR002182; NB-ARC.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00931; NB-ARC; 1.
CC Pfam; PF00400; WD40; 14.
CC PRINTS; PR00320; GPROTEINERPT.
CC ProDom; PD000018; WD40; 1.
CC SMART; SM00320; WD40; 14.
CC PROSITE; PS00878; WD_REPEATS_1; 9.
CC PROSITE; PS50082; WD_REPEATS_2; 14.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.

FT REPEAT 55 93
FT REPEAT 640 679
FT REPEAT 682 721
FT REPEAT 724 763
FT REPEAT 766 807
FT REPEAT 809 849
FT REPEAT 850 889
FT REPEAT 892 931
FT REPEAT 934 975
FT REPEAT 976 1017
FT REPEAT 1019 1059
FT REPEAT 1060 1101
FT REPEAT 1103 1143
FT REPEAT 1144 1183
FT REPEAT 1186 1227
SQ SEQUENCE 1258 AA; 139513 MW; 45DF03B91170C451 CRC64;
Query Match 16.9%; Score 458.5; DB 1; Length 1258;

Best Local Similarity 32.0%; Pred. No. 1.9e-25;
Matches 105; Conservative 43; Mismatches 101; Indels 79; Gaps 8;

QY 165 LRGESEVFCANVPVSDLLASGSDSTAIRIWLSENSTGSLVLRHCIRGGQVPS 224
Db 974 LEGHTDFIYGFDPDSQTLASASTDSSVRLMI---STGQCFCILLEH----- 1019
QY 225 NKDVTSLDW-----NSEGTLATGSYDGFARIWT-KDGNLASTLQGHKGPPIALKXNKK 277
Db 1020 -----TDWYAVVVFHPOGKIATGSADCTVKNLWISTGQCLTSLSEHSDKILGMAWSPD 1073
QY 278 GNFILSACVDKTIILDAHTGEAKQOPFHSAPALVDWQSNNTFASCSTDIMCIHVCKLG 337
Db 1074 GQLLASASADQSVRLWD-----CCTGRCVGI----- 1099
QY 338 QDRPIKTPQGHTEVNAIKWDPGTGNLLASGDDMTLKIWSMKODNCVHDIQQHKEIYTI 397
Db 1100 -----LRGHSNRYSAISPNGEIIATCTDTQTVKINDWQGGCKLKTLTGHTNWVFDI 1152
QY 398 KWSPTGPGTNNPNNMLASASFDSTVRLWDVDRGICHTLTKHQBPVYSVAFSPDGYL 457
Db 1153 AFSPDGK-----ILASASHDQTVRIWDVNTGRKHICHTHTLVSSVAFSPDGEVY 1203
QY 458 ASGSFDFKCVHWN--TOVCLHYLNGQVL 483
Db 1204 ASSGQDQTVRIWVYKTCGLQLLAKEL 1231

RESULT 8

YL24 ANASP
ID YL24 ANASP STANDARD; PRT; 1683 AA.
AC Q8YV57;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein all2124.
GN ALL2124.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.,
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: Contains 13 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003588; BAB73823.1;
CC PIR; AF2071; AF2071.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 14.
CC PRINTS; PR00320; GPROTEINERPT.
CC SMART; SM00320; WD40; 14.
CC PROSITE; PS00878; WD_REPEATS_1; 3.
CC PROSITE; PS50082; WD_REPEATS_2; 14.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1070 1109
FT REPEAT 1111 1150
FT REPEAT 1151 1150


```

RESULT 10
Y163 SYNY3
ID Y163 SYNY3 STANDARD; PRT; 1693 AA.
AC Q55563;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical WD-repeat protein sil0163.
GN SIL0163.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]_
SEQUENCE FROM N.A.
RP MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -!- SIMILARITY: Contains 16 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D63999; BAA10064.1; -.
CC PIR; S76086; S76086.
CC InterPro; IPR008941; TPR-like.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 16.
CC PRINTS; PR00320; GPROTEINRPT.
CC Prodom; PD000018; WD40; 13.
CC SMART; SM00320; WD40; 16.
CC PROSITE; PS00678; WD REPEATS 1; 8.
CC PROSITE; PS50082; WD REPEATS 2; 15.
CC PROSITE; PS50294; WD REPEATS REGION; 1.
CC KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1008 1042 WD 1.
FT REPEAT 1053 1083 WD 2.
FT REPEAT 1094 1124 WD 3.
FT REPEAT 1135 1165 WD 4.
FT REPEAT 1176 1206 WD 5.
FT REPEAT 1217 1247 WD 6.
FT REPEAT 1258 1288 WD 7.
FT REPEAT 1299 1329 WD 8.
FT REPEAT 1340 1370 WD 9.
FT REPEAT 1381 1411 WD 10.
FT REPEAT 1422 1452 WD 11.
FT REPEAT 1463 1493 WD 12.
FT REPEAT 1504 1534 WD 13.
FT REPEAT 1545 1575 WD 14.
FT REPEAT 1586 1616 WD 15.
FT REPEAT 1627 1657 WD 16.
FT REPEAT 1693 AA; 189935 MW; 0977A827A0251CFF CRC64;
Query Match 14.5%; Score 394.5; DB 1; Length 1693;
Best Local Similarity 28.2%; Pred. No. 1.3e-20;
Matches 115; Conservative 55; Mismatches 149; Indels 89; Gaps 12;
100 RDKLAQQQAAAAAASQSSAKKXGENTANGE--ENGAHTIANNHTDMVEVDGVEI 157
1362 RDKTARLTWTTEGCEVAVLAHQGVNREGQSPDQGVITGS---AKTQAQNNVLG---- 1414
158 PPNKAVLVGHSEVFICANNPNSDILASGSGSTARIWNLSENSTSGTQVLVLRHCIRE 217
1415 --KKLTVLGRHQDAVLNVRFPSPDSQYIVTASDKGTARVWN-----NTGRELAVLRH---- 1463

```

```

218 GQDVPFNKDVTSLDWNSBEGTLLATGSDYGFARITWKDGNLASTLQGHKGPFPALKWNKK 277
1464 -----YEKNIFAAPFSADQGFIVTASDDNTAGIWEIVGREGVICRGHSGPVVFAQFSAD 1517
278 GNFIISAGVYKTIIDAHTGEAKQFPFHSAPALDWDQSNNTFASCTDWCIIHVCKLG 337
1518 SRYILTASVDNTARIWDF-----LJ 1537
338 QDRPIKTFQGHTEVNAIKMDPTGNLIASCSDDMTLKIMSKQDNVHDLQOHNKEIYIT 397
1538 --RPLTLTLAGHSIVYQARPSPEGNLIATVADHTARLWD-RSGKTAVAVLYGHQGLGVIV 1594
398 KWSPTGPGCTNNPNANILMASASFDSTVRLWDVDRGCIHITLKHQEPYVSVAFSPPGRVL 457
1595 DWSPPGQ-----MLVTASNDGTARLWDLG-SRELLTLEGHGNVRSASFSPPGRVW 1644
458 ASGSFDCVCHWNTQVCLHYLN--GOVLLNLGRSICLYTIPHELHVIP 503
1645 LTSSADGTAKLPVKLTPLLQSGGQWLKXY-----LTHNALVSP 1684

RESULT 11
HUS7 HUMAN
ID HUS7 HUMAN STANDARD; PRT; 485 AA.
AC Q9NVX2; O60868; Q9BU54;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE WD-repeat protein HUSY-07.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
SEQUENCE FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]_
SEQUENCE FROM N.A.
RA TISSUE=Lung, and Placenta;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
EL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]_
SEQUENCE OF 400-485 FROM N.A.
RT TISSUE=Brain;
RX MEDLINE=21064499; PubMed=11124703;
RA Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,

```

```
RA Cannata N., Zimbello R., Lanfranchi G., Valle G.;
RT "Characterization of 16 novel human genes showing high similarity to
RL yeast sequences.";
RN Yeast 18:69-80(2001).
RN [4]
RX SUBCELLULAR LOCATION.
RP MEDLINE=22317277; PubMed=12429849;
RA Scherl A., Couté Y., Deon C., Calle A., Kindbeiter K., Sanchez J.-C.,
RT Greco A., Hochstrasser D.F., Diaz J.-J.;
RL "Functional proteomic analysis of human nucleolus.";
RN Mol. Biol. Cell 13:4100-4109(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK001320; BA91621.1; -
DR EMBL; BC002884; AA02884.2; -
DR EMBL; BC012075; AA012075.1; -
DR EMBL; AJ005257; CRA06444.1; -
DR SWISS-2DPAGE; Q9NVX2; HUMAN.
DR InterPro; IPR001632; Gprtein_B.
DR Pfam; PF00400; WD40.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS0082; WD REPEATS 2; 7.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Nuclear protein; Repeat; WD repeat.
FT REPEAT 112 151
FT REPEAT 154 193
FT REPEAT 197 241
FT REPEAT 244 282
FT REPEAT 325 366
FT REPEAT 370 409
FT REPEAT 412 451
FT REPEAT 454 484
SQ SEQUENCE 485 AA; 53266 MW; 0ABE24B44957379 CRC64;
Query Match
Best Local Similarity 13.5%; Score 366.5; DB 1; Length 485;
Matches 103; Conservative 47; Mismatches 145; Indels 89; Gaps 9;
QY 165 LRHSEVEFICAWNPVSDLLASGSDSTARIWNLSEN----STGSGTQLVLRHCIREGGQ 220
DB 110 LECHSEAVISVAFSPGKYLASGSGDTTVRFWDLSTETPHFTCKGHRHWLSISWSPDGK 169
QY 221 DYPSS-----NKDVTSLDW-----NSEGITLLATGSDVDGAR 250
DB 170 KLASGCKGQILLDPSTGKQVRLTAGHSKWITGLSWEPLHANPECRYVASSSKDGSVR 229
QY 251 IW-TKDGNLASTLGQHGKGFIFALKVKKGNFILTASGVDTTIIDAHGTGEAKQPPFH-- 307
DB 230 IWDTTAGRCERLTGHTQSVTCRLRGSDG-LLYSASQDRITIKVWAHDGVLCRTLOGHG 288
QY 308 --SAPALVDW-----
DB 289 WYNTWALSTDYALRTGAFEPAEASVNPDLQSLQELKERALSRYNLVRGQGPRLVSGS 348
QY 327 TDMCHVCKLQGD-RPIKTFQGHTEVNAIKWDPTGNILLASGDDMTLKWSMKQDCNVH 385
DB 349 DDFTLFLNSPAEDKPLRTWTCGQALINQVLFSPDSRVASASFDKSLKLDGRTGKYL 408
QY 386 DLQQHNKEIYTIKWSPTGFGTNNPNANMLASASFDSTVRLWDVDRGICHTLTGQEPV 445
```

```
DB 409 SURGHVAAYVQIWSAD-----SRLLVSGSDSTIKVWDVKAQKIAMDLPGHADEV 459
QY 446 YSVAFSPDGRYLASGFDKCVHIW 469
DB 460 YAVDWSPDGQRVASGKDKCLRIV 483
RESULT 12
HUS7.MOUSE
ID HUS7.MOUSE STANDARD; PRT; 485 AA.
AC Q8VEJ4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE WD-repeat protein HUSY-07.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Maman A., Rodriguez S., Sanchez A.,
RA Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; BC018399; AA018399.1; -
DR InterPro; IPR001632; Gprtein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS0082; WD REPEATS 2; 7.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Nuclear protein; Repeat; WD repeat.
FT REPEAT 112 151
FT REPEAT 154 193
FT REPEAT 197 241
FT REPEAT 244 282
FT REPEAT 327 366
FT REPEAT 370 409
```


FT REPEAT 412 451 WD 7.
 FT REPEAT 454 484 WD 8.
 SQ SEQUENCE 485 AA; 53148 MH; 2D7F59D603AEC07B CRC64;
 Query Match 13.4%; Score 365.5; DB 1; Length 485;
 Best Local Similarity 26.6%; Pred. No. 3.2e-19;
 Matches 102; Conservative 48; Mismatches 145; Indels 89; Gaps 9;
 QY 165 LRGESEVFICAWNPVSDLLASGSDSTARINWLSN-----STSGSTQLVLRHCIREGQ 220
 DB 110 LEHSEAVISVAFSPGKYLASGSDTTRFDLSTETPHFTCKGRRHWLVSISKSPDGK 169
 QY 221 DVPS-----NKDVTSLDW-----NSETLLATGSDVGFAR 250
 DB 170 KLASGCKNGQILLWDPSTGLQVTRTLTGHSKMITGLSWEPLHNPCECRYVASSRDXGYSR 229
 QY 251 IW-TKGNLASTLGOHQIFALKWKKGNFILSAGVDKTTIWDHAHTGEAKQPFPH-- 307
 DB 230 VMDITAGRCERILTGHSTQVCLRWGGDGLLYSASGRTIKVRAHDVLCRTLQGHCH 288
 QY 308 --SAPALVDW-----QSNNTFASCS 326
 DB 289 WNTMALSTIDYALRTGAFEPAEATVNAQDLQGLSKELKERASSRYNLVRGQGPRLVSGS 348
 QY 327 TDCIHKVCKGQD-RIKTFQGTNEVNAIKWDPTGNILLASCSDDMTLKIWSKQDNCVH 385
 DB 349 DDFTLFWSPAEDCKPLARMTGHQALINQVLFSPDSRIVASASFDSIKLMDGRTGKYLA 408
 QY 386 DLQHNKEIYTIKWSPTGPTNNPNANMLASAFDSTVRLMDVDRGICHTLTHQEPV 445
 DB 409 SLRSHVAATQIWSAD-----SRLVSGSSDSTLKVWDYKAQKLATDLPGHADV 459
 QY 446 YSAFSPDGRLASGSDPKCVHIW 469
 DB 460 YAVDWSPDQGVASGDKCLRIW 483
 RESULT 13
 LIS1 HUMAN STANDARD; PRT; 409 AA.
 ID AC P43034; Q8WZ88; Q8WZ89;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Platelet-activating factor acetylhydrolase IB alpha subunit (PAF
 DE acetylhydrolase 45 kDa subunit) (PAF-AH 45 kDa subunit) (PAF-AH alpha
 DE (PAPAH1) OR PAPAH1 (Lissencephaly-1 protein) (LIS-1).
 GN GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Kidney;
 RX MEDLINE=93361119; PubMed=835785;
 RA Reiner O., Carrozzo R., Shen Y., Wehnert M., Faustiniella F.,
 RA Dobyns W.B., Caskey C.T., Ledbetter D.H.;
 RT "Isolation of a Miller-Dieker lissencephaly gene containing G protein
 RT beta-subunit-like repeats.";
 RN Nature 364:717-721(1993).
 [2]
 RP SEQUENCE FROM N.A., AND VARIANT LIS1 ARG-148.
 RX MEDLINE=97217774; PubMed=9063735;
 RA Lo Nigro C., Chong S.S., Smith A.C.M., Dobyns W.B., Carrozzo R.,
 RA Ledbetter D.H.;
 RT "Point mutations and an intragenic deletion in LIS1, the lissencephaly
 RT causative gene in isolated lissencephaly sequence and Miller-Dieker
 RT syndrome.";
 RN Hum. Mol. Genet. 6:157-164(1997).
 [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;

RA Zhao M.J., Xia S.L., Li T.P.;
 RT "High expression of the lissencephaly gene in hepatocarcinoma
 RT patients.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RX Feng Z., Zhang B., Peng X., Yuan J., Qiang B.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RX Pobo G., Han M., Wiemann S.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 [6]
 RP SUBCELLULAR LOCATION, AND INTERACTION WITH DYNEIN AND DYNACTIN.
 RX MEDLINE=20512525; PubMed=11056532;
 RA Faulkner N.E., Dujardin D.L., Tai C.-Y., Vaughan K.T., O'Connell C.B.,
 RA Wang Y.-L., Vallee R.B.;
 RT "A role for the lissencephaly gene LIS1 in mitosis and cytoplasmic
 RT dynein function.";
 RL Nat. Cell Biol. 2:784-791(2000).
 [7]
 RP INTERACTION WITH RSN; DYNEIN AND DYNACTIN.
 RX MEDLINE=21898370; PubMed=11889140;
 RA Tai C.-Y., Dujardin D.L., Faulkner N.E., Vallee R.B.;
 RT "Role of dynein, dynactin, and CLIP-170 interactions in LIS1
 RT kinetochore function.";
 RL J. Cell Biol. 156:959-968(2002).
 [8]
 RP SUBCELLULAR LOCATION, AND INTERACTION WITH RSN.
 RX MEDLINE=21938534; PubMed=11940666;
 RA Coquelle F.M., Caspi M., Cordelières F.P., Dompierre J.P.,
 RA Dujardin D.L., Koifman C., Martin P., Hoogenraad C.C., Akhmanova A.,
 RA Galjart N., De Mey J.R., Reiner O.;
 RT "LIS1, CLIP-170's key to the dynein/dynactin pathway.";
 RL Mol. Cell Biol. 22:3089-3102(2002).
 [9]
 RP VARIANT SBH PRO-168.
 RX MEDLINE=99371784; PubMed=10441340;
 RA Pilz D.T., Kuc J., Matsumoto N., Bodurtha J., Bernadi B.,
 RA Tassinari C.A., Dobyns W.B., Ledbetter D.H.;
 RT "Subcortical band heterotopia in rare affected males can be caused by
 RT missense mutations in DCX (XLIS) or LIS1.";
 RN Hum. Mol. Genet. 8:1757-1760(1999).
 [10]
 RP VARIANTS LIS1 SER-30; SER-161 AND HIS-316.
 RX MEDLINE=21394321; PubMed=11502906;
 RA Leventer R.J., Cardoso C., Ledbetter D.H., Dobyns W.B.;
 RT "LIS1 missense mutations cause milder lissencephaly phenotypes
 RT including a child with normal IQ.";
 RL Neurology 57:416-422(2001).
 CC -!- FUNCTION: Probably involved in nuclear migration during cell
 CC division. Participates in the process of neuronal cell
 CC differentiation or brain development, possibly by through its role
 CC in targeting cytoplasmic dynein to the microtubule plus ends
 CC thereby playing an essential role in dynein-mediated microtubule
 CC sliding. Non-catalytic subunit of an acetylhydrolase complex, a
 CC complex that inactivates platelet-activating factor (PAF) by
 CC removing the acetyl group at the SN-2 position.
 CC -!- SUBUNIT: Cytosolic PAF-AH IB is formed of three subunits of 45 kDa
 CC (alpha), 30 kDa (beta) and 29 kDa (gamma). The catalytic activity
 CC of the enzyme resides in the beta and gamma subunits, whereas the
 CC alpha subunit has regulatory activity. Trimer formation is not
 CC essential for the catalytic activity. Interacts with dynein and
 CC dynactin. Interacts with RSN.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; localizes to the plus end of
 CC microtubules.
 CC -!- TISSUE SPECIFICITY: Fairly ubiquitous expression in both the
 CC frontal and occipital areas of the brain.
 CC -!- DISEASE: Defects in PAPAH1 are the cause of classical
 CC lissencephaly-1 (LIS1) [MIM:607432]. LIS1 is a brain malformation

caused by abnormal neuronal migration at 9 to 13 weeks' gestation. Lissencephaly means 'smooth brain', a brain without convolutions or gyri, and consists in an absence (agyria) or a decrease (pachygyria) of surface convolutions associated with a disorganization of the clear neuronal lamination of normal six-layered cortex. It is characterized by an abnormally thick and poorly organized cortex with 4 primitive layers, diffuse neuronal heterotopia, enlarged and dysmorphic ventricles, and often hypoplasia of the corpus callosum.

-1- DISEASE: Defects in PAFAH1B1 are the cause of subcortical band heterotopia (SBH) [MIM:607432]. SBH is at the less severe end of the lissencephaly spectrum of malformations, characterized by bilateral and symmetric ribbons of gray matter found in the central white matter between the cortex and the ventricular surface.

-1- DISEASE: Defects in PAFAH1B1 are a cause of Miller-Dieker lissencephaly syndrome (MDLS) [MIM:247200]. MDLS is a form of lissencephaly associated with facial abnormalities.

-1- SIMILARITY: Contains 1 Lish domain.

-1- SIMILARITY: Contains 7 WD repeats.

-1- CAUTION: Ref.1 (AA02882) sequence differs from that shown due to a chimeric cDNA.

-1- CAUTION: Ref.5 sequence differs from that shown due to what seems to be the presence of intronic sequence in the cDNA.

-1- DATABASE: NAME=HotMolbase; NOTE=LISI entry;

WWW="http://bioinformatics.weizmann.ac.il/hotmolbase/entries/lisi.htm".

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: L13385; AAA02880.1; -
EMBL: L13386; AAA02881.1; -
EMBL: L13387; AAA02882.1; ALT_SEQ.

EMBL: U72342; AAC51111.1; -

EMBL: U72334; AAC51111.1; JOINED.

EMBL: U72335; AAC51111.1; JOINED.

EMBL: U72336; AAC51111.1; JOINED.

EMBL: U72337; AAC51111.1; JOINED.

EMBL: U72338; AAC51111.1; JOINED.

EMBL: U72339; AAC51111.1; JOINED.

EMBL: U72340; AAC51111.1; JOINED.

EMBL: U72341; AAC51111.1; JOINED.

EMBL: AF208837; AAL34972.1; -

EMBL: AF208838; AAL34973.1; -

EMBL: AF400434; AAK92483.1; -

EMBL: BX538346; CAD98141.1; -

Genew; HGNC:8574; PAFAH1B1.

MIM: 601545; -

MIM: 607432; -

MIM: 247200; -

GO: GO:0006928; P:cell motility; TAS.

GO: GO:0006629; P:lipid metabolism; TAS.

GO: GO:0007399; P:neurogenesis; TAS.

GO: GO:0007165; P:signal transduction; TAS.

InterPro: IPR006594; Lish.

InterPro: IPR001680; WD40.

Pfam: PF00400; WD40; 7.

PRINTS: PR00320; GPROTEINBRPT.

ProDom: PD000018; WD40; 5.

SMART: SM00667; Lish; 1.

SMART: SM00320; WD40; 7.

PROSITE: PS00896; LISH; 1.

PROSITE: PS00678; WD_REPEATS_1; 4.

PROSITE: PS00882; WD_REPEATS_2; 7.

PROSITE: PS00294; WD_REPEATS_REGION; 1.

Cell division; Mitosis; Neurogenesis; Cytoskeleton; Microtubule;

Repeat; Coiled coil; WD repeat; Disease mutation; Lissencephaly.

INIT_MET 0 BY SIMILARITY.

FT

FT	DOMAIN	6	38	LISH.
FT	DOMAIN	50	81	COILED COIL (POTENTIAL).
FT	REPEAT	105	135	WD 1.
FT	REPEAT	147	177	WD 2.
FT	REPEAT	189	219	WD 3.
FT	REPEAT	231	261	WD 4.
FT	REPEAT	273	323	WD 5.
FT	REPEAT	335	365	WD 6.
FT	REPEAT	377	407	WD 7.
FT	VARIANT	30	30	F -> S (in LISI).
FT	VARIANT	148	148	/FTid=VAR_015398.
FT	VARIANT	161	161	H -> R (in LISI).
FT	VARIANT	168	168	/FTid=VAR_007724.
FT	VARIANT	168	168	G -> S (in LISI).
FT	VARIANT	168	168	/FTid=VAR_015399.
FT	VARIANT	316	316	S -> P (in SBH).
FT	VARIANT	316	316	/FTid=VAR_010203.
FT	VARIANT	20	20	D -> H (in LISI).
FT	CONFLICT	20	20	/FTid=VAR_015400.
FT	CONFLICT	20	20	S -> P (IN REF. 3).

Query Match 12.9%; Score 351.5; DB 1; Length 409;
Best Local Similarity 22.2%; Pred. No. 2.6e-18;
Matches 109; Conservative 77; Mismatches 165; Indels 141; Gaps 15;

QY	6	DEVNLFVRYLOESGFHSHTFTGKSHISOSNINGALVPPAALISITQKGLQVVEAEVS	65
QY	6	DEVNLFVRYLOESGFHSHTFTGKSHISOSNINGALVPPAALISITQKGLQVVEAEVS	65
DB	8	DELNRAIADYLRNGYEAE	37
QY	66	INEDGTLFDGPIESLSLIDAVMPDVVQTRQAVRDQLAQOQAAAAAASQGSAAK	125
DB	38	VNEE---LDKK---YAGLEKKWTSVIRLQKVVLESKLEAKBEFTSGGFLQKQDPK	91
QY	126	NGENTANGEANGAHTIANNHDTMMEVDGVEIPPKNKAVVLEHSEVPICAWNPVSDLLA	185
DB	92	EW-----IPRPPEK-YALSCHRSPTVTRVIFHEVFSVMV	123
QY	186	SGSGDSTARINLS-----ENSTSGTQLVLEHCTREGQDVPSNKDVTSLDWNSEGILLA	241
DB	124	SASEDATIKVWDYETGDFERTLKHTDSV-----QDI-----SFDHSGKLLA	165
QY	242	TGSYDGFARIWKDG-NLASTLGHQKGFIFALKWKNKGNFTLSAGVDTKTIINDAHTGEA	300
DB	166	SCSADMTIKLWDFQFECIRTHGHENSVVAIMPNGDHIVASRDKTIKWEVQIG--	223
QY	301	KQPFPHSAPALVDWQSNFTFSCSDMCIHVCKLQCDRPIKTFQGHTEVNAIKWDPT	360
DB	224	-----YC-----VKTFTGHREWVRVRENQD	244
QY	361	GNLLASGDDMTLKIWSMKQNCVHDLQOHNKETITIKWSP-----TGPGT--NNP	409
DB	245	GTLLASCSNDQTVRVVVAKECAELREHVVVEICISWAFESSYSSISATGSETKKS	304
QY	410	NANMLASAFDSTVRLWDVDRGICHTLTKHQPVEVYVAFSPDGRYLAGSFDKCVHIV	469
DB	305	KPGFLLSGSRDKTIKWDVSTGCMCLMTLVGHNDVNRVGLFHSKGKFLSCADDKTLRV	364
QY	470	N--TQVCLHYLN	479
DB	365	DYKWRGCKNTLN	376

RESULT 14
LISI_BOVIN
ID LISI_BOVIN STANDARD; PRT; 409 AA.
AC P43033;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Platelet-activating factor acetylhydrolase IB alpha subunit (PAF
DE acetylhydrolase 45 kDa subunit) (PAF-AH 45 kDa subunit) (PAF-AH alpha)
DE (PAF-AH alpha) (Lissencephaly-1 protein) (LISI-1).
GN PAFAH1B1 OR PAF-AH OR LISI.


```

Db 100 -----TLEGHENEKVSAWAPSGNLLATCSRDKSVWVWEDEDEYECVSLNSHTQ 151
QY 393 EIVTIKWSPTGCTNNPNANIMLASASDSTVRLW--DVDRGICIHITLTHQEPVVSVAF 450
Db 152 DVKHVVHP-----SQELLASAYDDTVKUYREEDDWVCATLEGHSTWWSLAF 202
QY 451 SPDGRYLASGSPDKCVHIWNTQVCILHYLNG 480
Db 203 DFIGQLASCSDDRIVRWR-----QYLPG 227

Search completed: August 9, 2004, 16:46:12
Job time : 10.6667 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:41:13 ; Search time 34.6667 Seconds
(without alignments)
4678.164 Million cell updates/sec

Title: US-09-987-701-4

Perfect score: 2718

Sequence: 1 MSISDEVNFLVRYLQESG.....LPHLWVFLVALIELLVK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mnc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2468	90.8	514	11	Q8BHJ5	Q8bhj5 mus musculus
2	2465	90.7	514	11	Q8CBG4	Q8cbg4 mus musculus
3	2464	90.7	514	11	Q9EQD4	Q9eqd4 mus musculus
4	2421.5	89.1	519	13	Q7SZM9	Q7szm9 xenopus lae
5	2273.5	83.6	527	11	Q8BYO4	Q8byo4 mus musculus
6	2269.5	83.5	527	11	Q8BMM0	Q8bmm0 mus musculus
7	2229	82.0	577	4	Q8GUV2	Q8guy2 homo sapien
8	1761.5	64.8	412	11	Q8COA1	Q8coai mus musculus
9	1715.5	63.1	700	5	Q9XZK1	Q9xzk1 drosophila
10	1251	46.0	513	10	Q9FN19	Q9fn19 arabidopsis
11	923.5	34.0	524	5	Q9SRJ9	Q9srj9 drosophila
12	882	32.5	201	11	Q8VEG3	Q8veg3 mus musculus
13	444.5	16.4	1227	16	Q8ZOR1	Q8zor1 anabaena sp
14	442	16.3	1356	3	Q8X1P3	Q8xlp3 podospora a
15	442	16.3	1356	3	Q8X1P5	Q8xlp5 podospora a
16	433.5	15.9	1356	3	Q8X1P4	Q8xlp4 podospora a

17	429.5	15.8	1189	16	Q8YL09	Q8yl09 anabaena sp
18	426	15.7	1708	16	Q8YZI2	Q8yzi2 anabaena sp
19	423.5	15.6	1376	3	Q8X1P2	Q8xlp2 podospora a
20	408	15.0	1711	16	Q8Z019	Q8z019 anabaena sp
21	404.5	14.9	934	16	Q8YZ23	Q8yz23 anabaena sp
22	395	14.5	1747	16	Q8Z020	Q8z020 anabaena sp
23	382.5	14.1	564	3	Q74845	Q74845 schizosacch
24	381	14.0	1551	16	Q8YMU3	Q8ymu3 anabaena sp
25	376	13.8	415	4	Q8N136	Q8n136 homo sapien
26	370.5	13.6	411	5	Q96698	Q96698 drosophila
27	369	13.6	676	16	Q8YSG6	Q8ysg6 anabaena sp
28	368.5	13.6	304	11	Q8DA72	Q8d4t2 mus musculu
29	368.5	13.6	304	16	Q8Z0F4	Q8z0f4 anabaena sp
30	368	13.5	415	4	Q8N776	Q8n776 homo sapien
31	367	13.5	410	13	Q803D2	Q803d2 brachydanio
32	366	13.5	502	3	Q74855	Q74855 schizosacch
33	365	13.4	1189	16	Q8YTD1	Q8ytd1 anabaena sp
34	360.5	13.3	410	13	Q902L4	Q902l4 xenopus lae
35	357.5	13.2	1241	2	Q9XB08	Q9xb08 amycolatops
36	356	13.1	352	10	Q80990	Q80990 arabidopsis
37	354	13.0	265	16	Q8YZ16	Q8yzi6 anabaena sp
38	353.5	13.0	476	13	Q93531	Q93531 xenopus lae
39	353.5	13.0	476	13	Q7ZXK9	Q7zxk9 xenopus lae
40	352	13.0	339	11	Q9DCZ7	Q9dcz7 mus musculu
41	352	13.0	339	11	Q99KN2	Q99kn2 mus musculu
42	351.5	12.9	1430	16	Q98HK1	Q98hk1 rhizobium l
43	349.5	12.9	410	6	Q9GL51	Q9gl51 sus scrofa
44	349.5	12.9	410	13	Q9PTR5	Q9ptr5 gallus gall
45	344.5	12.7	559	16	Q8YSC0	Q8ysc0 anabaena sp

ALIGNMENTS

RESULT 1

Q8BHJ5 PRELIMINARY; PRT; 514 AA.
AC Q8BHJ5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE IRAI protein.
GN IRAI OR 8030499H02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT Nature 420:563-573 (2002)."
RL 60,770 full-length cDNAs."
DR EMBL; AK029595; BAC26526.1; -
DR EMBL; AK033347; BAC28241.1; -
DR PIR; PT0651; PT0651.
DR MGD; MGI:2441730; Ira1.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR PRODOM; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS0082; WD REPEATS 2; 6.
DR PROSITE; PS00294; WD REPEATS REGION; 1.
SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F8BF14 CRC64;

Query Match 90.8%; Score 2468; DB 11; Length 514;
 Best Local Similarity 98.5%; Pred. No. 5e-167; Indels 0; Gaps 0;
 Matches 465; Conservative 2; Mismatches 5;

QY 1 MSISDEVNLFVRYLOBSGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
 DB 1 MSISDEVNLFVRYLOBSGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQAAAAAASQ 120
 DB 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQAAAAAASQ 120

QY 121 QGSAXNGENTANGEANGAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICAWNPV 180
 DB 121 QGSAXNGENTANGEANGAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICAWNPV 180

QY 181 SLLASGSGDSTARIMNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
 DB 181 SLLASGSGDSTARIMNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240

QY 241 ATGSYDGFARITWKDGNLASTLGQHKGPFPALKWNKGNFILSAGVDKTTIWDHTGEA 300
 DB 241 ATGSYDGFARITWKDGNLASTLGQHKGPFPALKWNKGNFILSAGVDKTTIWDHTGEA 300

QY 301 KQPFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQHTNEVNAIKWDPT 360
 DB 301 KQPFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQHTNEVNAIKWDPT 360

QY 361 GNLLASCSDDMTLKIKSMKQDNCVHDLQAHNKETIYIKWSPGPTGNNPNNANMLASAF 420
 DB 361 GNLLASCSDDMTLKIKSMKQDNCVHDLQAHNKETIYIKWSPGPTGNNPNNANMLASAF 420

QY 421 DSTVRLWDVDRGICHTLTQKQEPVSVAFSPDGRYLASGDFKCVHIWNTQ 472
 DB 421 DSTVRLWDVDRGICHTLTQKQEPVSVAFSPDGRYLASGDFKCVHIWNTQ 472

RESULT 2

Q8CBG4 PRELIMINARY; PRT; 514 AA.

AC Q8CBG4;
 DT 01-VAR-2003 (TRENBLrel. 23, Created)
 DT 01-VAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE IRAL protein.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RC MEDLINE=22354583; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK036064; BAC29294.1;
 DR PIR; PT0651; PT0651.
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000019; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00896; Lish; 1.
 DR PROSITE; PS00678; WD REPEATS 1; 4.
 DR PROSITE; PS00082; WD REPEATS 2; 6.
 DR PROSITE; PS00294; WD REPEATS_REGION; 1.

SQ SEQUENCE 514 AA; 55689 MW; 13BED3753A725029 CRC64;
 Query Match 90.7%; Score 2465; DB 11; Length 514;
 Best Local Similarity 98.3%; Pred. No. 8.2e-167;
 Matches 464; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSISDEVNLFVRYLOBSGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
 DB 1 MSISDEVNLFVRYLOBSGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQAAAAAASQ 120
 DB 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQAAAAAASQ 120

QY 121 QGSAXNGENTANGEANGAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICAWNPV 180
 DB 121 QGSAXNGENTANGEANGAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICAWNPV 180

QY 181 SLLASGSGDSTARIMNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
 DB 181 SLLASGSGDSTARIMNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240

QY 241 ATGSYDGFARITWKDGNLASTLGQHKGPFPALKWNKGNFILSAGVDKTTIWDHTGEA 300
 DB 241 ATGSYDGFARITWKDGNLASTLGQHKGPFPALKWNKGNFILSAGVDKTTIWDHTGEA 300

QY 301 KQPFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQHTNEVNAIKWDPT 360
 DB 301 KQPFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQHTNEVNAIKWDPT 360

QY 361 GNLLASCSDDMTLKIKSMKQDNCVHDLQAHNKETIYIKWSPGPTGNNPNNANMLASAF 420
 DB 361 GNLLASCSDDMTLKIKSMKQDNCVHDLQAHNKETIYIKWSPGPTGNNPNNANMLASAF 420

QY 421 DSTVRLWDVDRGICHTLTQKQEPVSVAFSPDGRYLASGDFKCVHIWNTQ 472
 DB 421 DSTVRLWDVDRGICHTLTQKQEPVSVAFSPDGRYLASGDFKCVHIWNTQ 472

RESULT 3

Q3EQD4 PRELIMINARY; PRT; 514 AA.

AC Q3EQD4;
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE IRAL.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RA Zhang X.; Dormady S.; Basch R.;
 RT "Identification of four human cDNAs that are differentially expressed
 RT by early hematopoietic progenitors."
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF268195; AAG4738.1;
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000019; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00896; Lish; 1.
 DR PROSITE; PS00678; WD REPEATS 1; 4.
 DR PROSITE; PS00082; WD REPEATS 2; 6.
 DR PROSITE; PS00294; WD REPEATS_REGION; 1.
 KW Repeat; WD repeat.

SQ SEQUENCE 514 AA; 55689 MW; 6A72CE68A40C141F CRC64;

Query Match
Best Local Similarity 90.7%; Score 2464; DB 11; Length 514;
Matches 464; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSISSEVNFVLYRVLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSISSEVNFVLYRVLQESGFSHSAFTFGIESHSQSNINGALVPPAALISIIQKGLQYV 60
QY 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAAASQ 120
DB 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAAATNQ 120
QY 121 QGSAKNGENTANGEENGCAHTIANNHTDMVEVDGVEIPPNKAVILRGHSEVEFICAWNPV 180
DB 121 QGSAKNGENTANGEENGCAHTIANNHTDMVEVDGVEIPPNKAVILRGHSEVEFICAWNPV 180
QY 181 SLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPNSKDVTSLDWNSBCTLL 240
DB 181 SLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPNSKDVTSLDWNSBCTLL 240
QY 241 ATGSYDGFARITWKDGNLSTLQHGKGPFPALKWKNKGFIILSAGVDKTTIIWDAHTGEA 300
DB 241 ATGSYDGFARITWKDGNLSTLQHGKGPFPALKWKNKGFIILSAGVDKTTIIWDAHTGEA 300
QY 301 KQGFPHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
DB 301 KQGFPHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASASF 420
DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASASF 420
QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQ 472
DB 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQ 472

RESULT 4
Q7SZM9 PRELIMINARY; PRT; 519 AA.
AC Q7SZM9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Nuclear receptor co-repressor complex subunit TBLR1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12794076;
RA Tomita A., Buchholz D.R., Obata K., Shi Y.B.;
RT "Fusion Protein of Retinoic Acid Receptor {alpha} with Promyelocytic Leukemia Protein or Promyelocytic Leukemia Zinc Finger Protein Recruits N-CoR-TBLR1 Corepressor Complex to Repress Transcription in Vivo.";
RL J. Biol. Chem. 278:30788-30795(2003).
DR EMBL; AY225088; AAP20646.1; -;
KW Receptor.
SQ SEQUENCE 519 AA; 55043 MW; 5E998EDC8C92296 CRC64;

Query Match
Best Local Similarity 89.1%; Score 2421.5; DB 13; Length 519;
Matches 458; Conservative 4; Mismatches 10; Indels 5; Gaps 1;

QY 1 MSISSEVNFVLYRVLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSISSEVNFVLYRVLQESGFSHSAFTFGIESHSQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAA-- 118
DB 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAA 120
QY 119 ---SQGSAKNGENTANGEENGCAHTIANNHTDMVEVDGVEIPPNKAVILRGHSEVEFIC 175
DB 121 TPNQOPPAKNGENTANGEENGCAHTIANNHTDMVEVDGVEIPPNKAVILRGHSEVEFIC 180
QY 176 AWPVSDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPNSKDVTSLDWNS 235
DB 181 AWPVSDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPNSKDVTSLDWNS 240
QY 236 EGTLLATSGYDGFARITWKDGNLSTLQHGKGPFPALKWKNKGFIILSAGVDKTTIIWDA 295
DB 241 EGTLLATSGYDGFARITWKDGNLSTLQHGKGPFPALKWKNKGFIILSAGVDKTTIIWDA 300
QY 296 HTGEAKQGFPHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAI 355
DB 301 HTGEAKQGFPHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAI 360
QY 356 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANML 415
DB 361 KWDPTGNLLASCSDDMTLKIWSMKHDTICVHDLQAHNKEIYTIKWSPTGPGTNNPNANML 420
QY 416 ASASFDSVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQ 472
DB 421 ASASFDSVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQ 477

RESULT 5
Q8BYQ4 PRELIMINARY; PRT; 527 AA.
AC Q8BYQ4
DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transducin (Tblix protein).
GN TBL1X.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hypothalamus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Joqueillano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalka U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

```

RT and mouse cDNA sequences." ;
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3].
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RA Strausberg R.; submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK038674; BAC30092.1; -.
DR EMBL; BC043105; AAH43105.1; -.
DR PIR; PT0651; PT0651.
DR MGI; MGI:1336172; Tbl1x.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
SQ SEQUENCE 527 AA; 56802 MW; 146435A9C51DFDA6 CRC64;

Query Match 83.6%; Score 2273.5; DB 11; Length 527;
Best Local Similarity 87.4%; Pred. No. 3.4e-153;
Matches 425; Conservative 26; Mismatches 20; Indels 15; Gaps 2;

QY 1 MSISDSEVNFVLYRLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSITSDEVNFVLYRLOESGFSHSAFTFGIESHSQSNINGTLVPPAALISIIQKGLQYV 60
QY 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVOTROQAFREKLAQOQAAAAA----- 115
DB 61 EAEISINEDGTVFDGRPIESLSLIDAVMPDVVOTROQAFREKLAQOQAAAAAATAAT 120
QY 116 -----AAASQGSANGENTANGENGAAHTIANNHTDMMEVDGVEIIPPNAKAVL 166
DB 121 ATSTAATPAAQAQNPKNGEATVNGEENGAAHAI -NNHSPKPEIDGVEIIPPNAKAVL 179
QY 167 GHESEVFICAWNPNVSDLLASGSDSTARIWNLSNSTSGTQVLRHCIREGGQDVPSNK 226
DB 180 GHESEVFICAWNPNVSDLLASGSDSTARIWNLSNSTSGTQVLRHCIREGGHDFPSNK 239
QY 227 DVTSLDWNSDGLTATGSDYDGFARITWKDGNLASTLGQHGKPIFALKWNKKNFILSAGV 286
DB 240 DVTSLDWNSDGLTATGSDYDGFARITWKDGNLASTLGQHGKPIFALKWNKKNFILSAGV 299
QY 287 DKTTIWDAAHTGEAKQOFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 346
DB 300 DKTTIWDAAHTGEAKQOFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPVKTFQ 359
QY 347 GHTNEVNAIKWDPPTGNILLASCDSDMTLKIWSMKQDNCVHDLQOHKEIYTIKNSPTGPT 406
DB 360 GHTNEVNAIKWDPGMLLASCDSDMTLKIWSMKQDACVHDLQAHKSEIYTIKNSPTGPT 419
QY 407 NNPANMLASASFDSTVRLWDVDRGICITHTLTKHQEPVYVAFSPDGRYLASGDFKCV 466
DB 420 SNPSNIMLASASFDSTVRLWDVVERGVCITHTLTKHQEPVYVAFSPDGRYLASGDFKCV 479
QY 467 HIWNTQ 472
DB 480 HIWNTQ 485

RESULT 6
Q8BMMO PRELIMINARY; PRT; 527 AA.
ID Q8BMMO
AC Q8BMMO;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Transducin.

GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pituitary;
MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs." ;
DR Nature 420:563-573 (2002).
DR EMBL; AK030547; BAC27015.1; -.
DR PIR; PT0651; PT0651.
DR MGI; MGI:1336172; Tbl1x.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
SQ SEQUENCE 527 AA; 56832 MW; 0466F764ABA25C80 CRC64;

Query Match 83.5%; Score 2269.5; DB 11; Length 527;
Best Local Similarity 87.2%; Pred. No. 6.6e-153;
Matches 424; Conservative 26; Mismatches 21; Indels 15; Gaps 2;

QY 1 MSISDSEVNFVLYRLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSITSDEVNFVLYRLOESGFSHSAFTFGIESHSQSNINGTLVPPAALISIIQKGLQYV 60
QY 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVOTROQAFREKLAQOQAAAAA----- 115
DB 61 EAEISINEDGTVFDGRPIESLSLIDAVMPDVVOTROQAFREKLAQOQAAAAAATAAT 120
QY 116 -----AAASQGSANGENTANGENGAAHTIANNHTDMMEVDGVEIIPPNAKAVL 166
DB 121 ATSTAATPAAQAQNPKNGEATVNGEENGAAHAI -NNHSPKPEIDGVEIIPPNAKAVL 179
QY 167 GHESEVFICAWNPNVSDLLASGSDSTARIWNLSNSTSGTQVLRHCIREGGQDVPSNK 226
DB 180 GHESEVFICAWNPNVSDLLASGSDSTARIWNLSNSTSGTQVLRHCIREGGHDFPSNK 239
QY 227 DVTSLDWNSDGLTATGSDYDGFARITWKDGNLASTLGQHGKPIFALKWNKKNFILSAGV 286
DB 240 DVTSLDWNSDGLTATGSDYDGFARITWKDGNLASTLGQHGKPIFALKWNKKNFILSAGV 299
QY 287 DKTTIWDAAHTGEAKQOFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 346
DB 300 DKTTIWDAAHTGEAKQOFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPVKTFQ 359
QY 347 GHTNEVNAIKWDPPTGNILLASCDSDMTLKIWSMKQDNCVHDLQOHKEIYTIKNSPTGPT 406
DB 360 GHTNEVNAIKWDPGMLLASCDSDMTLKIWSMKQDACVHDLQAHKSEIYTIKNSPTGPT 419
QY 407 NNPANMLASASFDSTVRLWDVDRGICITHTLTKHQEPVYVAFSPDGRYLASGDFKCV 466
DB 420 SNPSNIMLASASFDSTVRLWDVVERGVCITHTLTKHQEPVYVAFSPDGRYLASGDFKCV 479
QY 467 HIWNTQ 472
DB 480 HIWNTQ 485

RESULT 7
Q86UY2

```


ID AC Q86UY2 PRELIMINARY; PRT; 577 AA.
Q86UY2; 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE TBL1X protein.
DB Homo sapiens (Human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Altschuler R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin I.B., Toshlyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Rahak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Krzyszynski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052304; AAH52304.1;
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS50896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS50082; WD REPEATS 2; 6.
DR PROSITE; PS50294; WD REPEATS REGION; 1.
SQ SEQUENCE 577 AA; 62495 MW; D830A37781E2A15C CRC64;
Query Match 82.0%; Score 2229; DB 4; Length 577;
Best Local Similarity 85.4%; Pred. No. 5.6e-150;
Matches 414; Conservative 32; Mismatches 25; Indels 14; Gaps 2;
QY 1 MSISDEVNLFVYRLOESGFSGHSAFTFGIKSHISQSNINGALVPPAALISITQKLOYV 60
DB 52 MSITSDVNFVYRLOESGFSGHSAFTFGIESHISQSNINGTLVPPAALISITQKLOYV 111
QY 61 EAEVGINEDGTLFDGRPIESLSLIDAVMPDVVQVTRQAFREKLAQQQASAAAAATAA 171
DB 112 EAEISINEDGVFDGRPIESLSLIDAVMPDVVQVTRQAFREKLAQQQASAAAAATAA 171
QY 118 -----ASQGSAGKNGENTANGAHTIANNHTDMVEVDGDIPEIPNKAVLRG 167
DB 172 ATAATTSAGVSHQNPKNREATVNGEENRAHSV--NNHAKPWEIDGVEIPEIPSKATVLRG 230
QY 168 HESEVFICAMPVSDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQVPSNKD 227
DB 231 HESEVFICAMPVSDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQVPSNKD 290

QY 228 VTSLDWNSEGTLLATGSDYGFARIWTKDGNLASTLGOHKGPFPALKKWKNGNFIISAGVD 287
DB 291 VTSLDWNSEGTLLATGSDYGFARIWTKDGNLASTLGOHKGPFPALKKWKNGNFIISAGVD 350
QY 288 KTTIWDHAHTGEAKQOFPFHSAPALVDVQSNNTTASCSTDMCIHVCKLGQDRPKTFQG 347
DB 351 KTTIWDHAHTGEAKQOFPFHSAPALVDVQSNNTTASCSTDMCIHVCKLGQDRPKTFQG 410
QY 348 HTNEVNAIKWDPDTGNLLASCSDDMTLKWSMKQDNCVHDLOQHNKEIYTIKNSPTGPGTN 407
DB 411 HTNEVNAIKWDPDTGNLLASCSDDMTLKWSMKQDNCVHDLOQHNKEIYTIKNSPTGPGTN 470
QY 408 NPENAMLMASAFDSTVRLMDVDRGICHTLTKHQPVPVSVAFSPDGRYLASSGSPDKCVH 467
DB 471 NPENAMLMASAFDSTVRLMDVDRGICHTLTKHQPVPVSVAFSPDGRYLASSGSPDKCVH 530
QY 468 IWNTO 472
DB 531 IWNTO 535
RESULT 8
O8COA1 PRELIMINARY; PRT; 412 AA.
ID O8COA1
AC O8COA1;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Transducin (Fragment).
GN TBL1X.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
DR EMBL; AK031937; BAC27612.1; -.
DR FIRM; PT0651;
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS50082; WD REPEATS 2; 6.
DR PROSITE; PS50294; WD REPEATS REGION; 1.
FT NON TER 1
SQ SEQUENCE 412 AA; 44356 MW; EB78910E6D9E5237 CRC64;

Query Match 64.8%; Score 1761.5; DB 11; Length 412;
Best Local Similarity 87.0%; Pred. No. 5.8e-117;
Matches 321; Conservative 20; Mismatches 27; Indels 1; Gaps 1;
QY 104 AQQQAAAAAASQGSAGKNGENTANGAHTIANNHTDMVEVDGDIPEIPNKAV 163
DB 3 AATATSTAATTPAARAAQNPFPNGEATVNGEENRAHAI--NNHAKPWEIDGDIPEIPSKAT 61
QY 164 VLRGHSEVFICAMPVSDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQVDP 223
DB 62 VLRGHSEVFICAMPVSDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQVDP 121
QY 224 SNKDVTSLOWNSGTLATGSDYGFARIWTKDGNLASTLGOHKGPFPALKKWKNGNFIIS 283
DB 122 SNKDVTSLOWNSGTLATGSDYGFARIWTKDGNLASTLGOHKGPFPALKKWKNGNFIIS 181

QY 284 AGVDTKTIITWDHTGEAKQOPPHSAPALDWDWQSNNTFASCSTDMCIHVCKLQODRPIK 343
 Db 182 AGVDTKTIITWDHTGEAKQOPPHSAPALDWDWQSNNTFASCSTDMCIHVCKLQODRPIK 241
 QY 344 TFOGHNEVNAIKWPBTGNLLASCDDMTLTKWSMKQDNCVHDLQOHKNIYIKRSPG 403
 Db 242 TFOGHNEVNAIKWPBTGNLLASCDDMTLTKWSMKQDNCVHDLQOHKNIYIKRSPG 301
 QY 404 PGTNNPNAMLASASFTSTVLWLDVDRGICIHILTHKQBPVYVAFSPDGRVYLAGSFD 463
 Db 302 PATSNFNSNIMLASASFTSTVLWLDVDRGICIHILTHKQBPVYVAFSPDGRVYLAGSFD 361
 QY 464 KCVHWTNTQ 472
 Db 362 KCVHWTNTQ 370

RESULT 9
 Q9XZK1 PRELIMINARY; PRT: 700 AA.
 AC Q9XZK1; (TREMELrel. 12, Created)
 DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE EBI protein.
 GN EBI OR CG4063.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Fabios B., Delcher A., Deng Z., Mays A.B., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92234084; PubMed=10215623;
 RA Dong X., Tsuda L., Zavitz K.H., Lin M., Li S., Carthew R.W.,
 RA Zipursky S.L.;
 RT "ebi regulates epidermal growth factor receptor signaling pathways in
 RT Drosophila";
 RL Genes Dev. 13:954-965(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Celisner S.E., George R.A., Galle R.F., Hoskins R.A., Svitskas R.R.,
 RA Harris N.L., Agbayani A., Arcaina T.T., Baxter E., Blazey R.G.,
 RA Chavez C., Chew M., Doyle C.M., Farfan D.E., Flanagan J.,
 RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim S.H., Lee B.,
 RA Lomotan M.A., Mak J., Mazda P., Moshrefi A.R., Moshrefi W., Nixon K.,
 RA Pacle J.M., Park S., Pfeiffer B., Punch E., Snir E., Twomey B.,
 RA Wan K.H., Zhang R., Zieran L.L., Rubin G.M.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003589; AAF51501.1; -;
 DR EMBL; AF146345; AAD35017.1; -;
 DR FlyBase; FBgn0023444; cbi.
 DR CG; GO:0000074; P:regulation of cell cycle; IMP.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001005; MYB_DNA_binding.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000015; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00896; Lish; 1.
 DR PROSITE; PS00037; MYB_1; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 3.
 DR PROSITE; PS00082; WD_REPEATS_2; 6.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 700 AA; 72387 MW; 28C6D8D07B79FB7 CRC64;
 Query Match 63.1%; Score 1715.5; DB 5; Length 700;
 Best Local Similarity 52.9%; Pred. No. 2.3e-113;
 Matches 349; Conservative 49; Mismatches 71; Indels 191; Gaps 8;
 QY 1 MSISDENVFLVRYLQESGFHSFTFGIKSHISOSNINGALVPPAALISIQKGLV 60
 Db 1 MSFSSDENVFLVRYLQESGFHSFTFGIKSHISOSNINGALVPPAALISIQKGLV 60
 QY 61 EAEVSTINEDGTLFDGRPIESLSLDAMPDV-----VOTR----- 96
 Db 61 EAEVSTINEDGTLFDGRPIESLSLDAMPDV-----VOTR----- 96
 QY 61 EAEVSTINEDGTLFDGRPIESLSLDAMPDV-----VOTR----- 96
 Db 61 EAEVSTINEDGTLFDGRPIESLSLDAMPDV-----VOTR----- 96
 QY 97 -----QAYRDKLA-----QOAAAAAASOOSAKNGENT 130
 Db 97 -----QAYRDKLA-----QOAAAAAASOOSAKNGENT 130
 QY 119 NAKPEILIEPTGTGAGSAGNGKAGTGTGTTPTDQASAEVDSGNGNANNAGTAVGNG 178
 Db 119 NAKPEILIEPTGTGAGSAGNGKAGTGTGTTPTDQASAEVDSGNGNANNAGTAVGNG 178
 QY 131 ANGE----- 135
 Db 131 ANGE----- 135
 QY 179 AGNGAAGTGGSTNSTTSTPAGGLAAGPAGSAGKQSGNSAGSSSGNAGNANATSDAASS 238
 Db 179 AGNGAAGTGGSTNSTTSTPAGGLAAGPAGSAGKQSGNSAGSSSGNAGNANATSDAASS 238
 QY 136 -----NG-----AHTIANNHTM----- 148
 Db 136 -----NG-----AHTIANNHTM----- 148
 QY 239 TSTNGNSTSSVEQPTSGTGTGTTPTDQASAEVDSGNGNANNAGTAVGNG 298
 Db 239 TSTNGNSTSSVEQPTSGTGTGTTPTDQASAEVDSGNGNANNAGTAVGNG 298
 QY 149 -----MEVDGVEIPPKAVVLRGHESEV 172
 Db 149 -----MEVDGVEIPPKAVVLRGHESEV 172
 QY 299 GNNVQSSSSNAQSAFSGTISSTSGAGTGAALVPMIDENIEIPESKARVLRGHESEV 358
 Db 299 GNNVQSSSSNAQSAFSGTISSTSGAGTGAALVPMIDENIEIPESKARVLRGHESEV 358
 QY 173 FICAWNPVSDLLASGSGDSTARINLSNSTSGTQLVLRHCIRREGGQDVPSPKDVTSID 232
 Db 173 FICAWNPVSDLLASGSGDSTARINLSNSTSGTQLVLRHCIRREGGQDVPSPKDVTSID 232
 QY 359 FICAWNPVSDLLASGSGDSTARINLSNSTSGTQLVLRHCIRREGGQDVPSPKDVTSID 417
 Db 359 FICAWNPVSDLLASGSGDSTARINLSNSTSGTQLVLRHCIRREGGQDVPSPKDVTSID 417
 QY 233 WNSECTLATGSDYDGPARTWTKDGNLSTLQHKGPFPALKKNGKNTLSAGVDTKTI 292

```

Db 418 WNCDSLLATSGYGVARIWKTGRLATLQKGFIFALKWKNCGNYILSAGVDKTTII 477
QY 293 WDAHTGKQFPFHSAAPALVDWQSNUPFASCTDWCIVKCLGODRPIKIFOGHTNEV 352
Db 478 WDASTGCTQOFAHFAAPALVDWQINQAFASCTDQRHVCRLGWNPKIFKGTNEV 537
QY 353 NAIKWPTGNLLASCSDDMTLKIWSKQDNCVHDLOQHNKEIYTIKWSPTGPGTNNPNAN 412
Db 538 NAIKWCFQOGLLASCSDDMTLKIWSMNDRCCDLQAHSEIYTIKWSPTGPGTNNPN 597
QY 413 LNLASAFSDTVRLWVDRCICHTITLKQEPYVYAFSPDGRYLAGSFDKCVHIWNTQ 472
Db 598 LILASAFSDTVRLWVDVRSCHITLTKTEPYVYAFSPDGKHLASGFDKCVHIWSTQ 657

RESULT 10
Q9FN19 PRELIMINARY; PRT; 613 AA.
AC Q9FN19;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genomic DNA, chromosome 5, TAC clone:KaK14 (AT5g67320/KaK14_4).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asanizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned pl clones.";
RL DNA Res. 4:401-414(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.W., Chang C.H., Dale J.M., Deng J.M., Hayashizaki Y.,
RA Hsuan W.W., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M.,
RA Southwick A., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yamada K., Yu G., Yuan S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007645; BAB09017.1; -
DR EMBL; AY057698; AAL15328.1; -
DR EMBL; AY143932; AAN28871.1; -
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 2.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.

```

```

DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 613 AA; 69772 NW; 04F40772311COE76 CRC64;

Query Match
Best Local Similarity 46.0%; Score 1251; DB 10; Length 613;
Matches 248; Conservative 85; Mismatches 131; Indels 108; Gaps 7;

QY 2 SISDEVNLFVYRLOESGFSHSAFTFGIKSHISCSNINGALVPPAALISIQKGLQVVE 61
Db 3 SLTSVELNLFVRYLQESGFTAAFTLGEAGINKNSIDGNMVPFGALIKVQKGLQVME 62
QY 62 -----AEVSINEDGTLFDGRPIESLSLIDAVMPDVVTRQAVRDKLACQAAAAA 114
Db 63 MEANLSNEVIDEDFSFF--QPLDLISKVYKELQDMLREKKRKERDMEKEDSKENDK 120
QY 115 AAAAQSGSAGKNGENTANGE-----ENGAHTIAN 143
Db 121 GVERHEGDRNRAKEKDRHOKERERERERERERERERERERERERERERERER 180
QY 144 NHTDMWEVD-----GDVE----- 156
Db 181 REKORLKEKEEREIERERERERERERERERERERERERERERERERERERERER 240
QY 157 -----IPFNKAVILRGHSEVFI CAMNPVSDLLASGSDSTARIWNLSENSTS 204
Db 241 DIVMTPTSTSHIPNSDVRILGHTSEVCAWSPSALLASGSDATARIWISPEGSFK 300
QY 205 -----GSTQLVLRHCIREGQDVPSKDVTSIDWNSGTLATGSDGPARIWTKGNL 258
Db 301 AVHTGRTNAILLXHA---KGSNEKSGKDVTTLDWNGEGTLLATGSCDQARITWLGEL 357
QY 259 ASTLQGHKGPIFALKWKNKGNFILSAGVDKTTIWDATHTGEAKQFPFHSAAPALVDWQS 318
Db 358 1STLSKHKGPIFSLKWNKKGDIYLLTGSVDRTAVVDVKAEWKKQFPFHSGFTLDVDRN 417
QY 319 NNTFASCTDMCIHVCKLQDPRPKTFQGHNTNEVNAIKWDPTGNLLASCSDDMTLKWSM 378
Db 418 NVSFATSTDSMIYLCIGETRPARTFTGHQGVNVCVKWDPTGSLLASCSDDSTAKIWN 477
QY 379 KODNCVHDLQHNKEIYTIKWSPTGPGTNNPNANMLASAFSDTVRLWVDVRCICHTL 438
Db 478 KQSTFVHDLRHTKEIYTIKWSPTGPGTNNPNKQTLTASAFSDTVKLWDAELGKMLCSF 537
QY 439 TKHQEPYVYAFSPDGRYLAGSFDKCVHIWN 470
Db 538 NGHREPVYSLAFSPNGEYIAGSLDKSIHWS 569

RESULT 11
Q9SRJ9 PRELIMINARY; PRT; 524 AA.
ID Q9SRJ9;
AC Q9SRJ9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE LD24373p.
DE EBI OR CG4063.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brckstein P., Hong L., Agbavani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,

```



```

DR Pfam; PF00400; WD40; 14.
DR PRINTS; PR00364; DISEASERSIST.
DR PRINTS; PR00320; GPROTEINBRPT.
DR PRODOM; PD000018; WD40; 1.
DR SMART; SM00320; WD40; 14.
DR TIGRFAMs; TIGR01199; HTH fis; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 12.
DR PROSITE; PS50082; WD_REPEATS_2; 14.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Complete proteome.
SQ SEQUENCE 1227 AA; 137236 MW; 466F726939ED4FBF CRC64;

Query Match          16.4%; Score 444.5; DB 16; Length 1227;
Best Local Similarity 29.2%; Pred No. 9.9e-23;
Matches 130; Conservative 73; Mismatches 171; Indels 71; Gaps 15;

Qy 64 VSINEDG-TLFDGRIPIESLIDAVNPVQTRQ-----QAYRDKLAQQAAAAA 117
Db 780 VSFSPDQQTLLASGEDSTVRLWD-----VKTGQWQIFEGHSKKVSVRFSPDGQTLAS 833
Qy 118 ASQGSRAK-----NGE--NTANGSENGAHTIANNHTDMEVDG-----DV 155
Db 834 CGEDRSIKLWDIQEGECVNTLWGHSSQWAIA-----FSPDGRTLISCDDQTPARLWDV 887
Qy 156 EIPPNKAVLRGHESEYFICAWNVPVSDLLASGSGDSTARINWSENSTSGTQLVLRHCI 215
Db 888 -ITGNSLNLGRYTRDVSVAFPSDQILASGRDDYTIGLWNLKTGEC-----HPL 937
Qy 216 REGQDVPSKNDVTSLDWNSGTTLLATGSDYDGFARIW----TKDGNLASTLGQHKGFIFA 271
Db 938 R-QHGQ-----RINSVAFHPDGKILASGADNTIKLWDISDTHNSKYIRTLTGHTNWWVT 991
Qy 272 LKWNKGNFILSAGVDKTTIIWAHTGEAKQPPFHSAPALDVDWQ--SNNTFASCSTDMC 330
Db 992 VVSPDKHTLLASSEDTIRLWDKDTGDCIQLKGHSWWTVAFPSDQILASGSADSE 1051
Qy 331 IHVCKLGQDRPIKTFQGHTEVNAIKWDPTGNLIASCDMTLKIWSMKQDNCVHDLQQH 390
Db 1052 IKIWDVASGKCLQTLTPDQGMINSVAFSLDGTLLASASEDTVKLWNLKTGECVHTLKGH 1111
Qy 391 NKEIYTIKWSPGTGPTNPNANMLASASPDSTVRLWDVDRGICHTLTK-HQEPVYSVA 449
Db 1112 EKQVYSVAFSNGQ-----IAASGSEDTTVKLWDISTGSCVDTLKGHTAAIKRSA 1162
Qy 450 FSPDGRYLAGSFDKCVHIWNTQVC 474
Db 1163 FSPDGRLLASGSEDEKIQLDWMQNC 1187

```

```

RESULT 14
Q8X1P3
ID Q8X1P3 PRELIMINARY; PRT; 1356 AA.
AC Q8X1P3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta transducin-like protein HET-E2c*40.
GN HET-E.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Bequeret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of her-e and her-d gensa.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF523584; AAL37300.1; -.
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF05729; NACHT; 1.

```

DR	Pfam; PF00400; WD40; 10.
DR	PRINTS; PR00320; PROTEINERPT.
DR	ProDom; PD000018; WD40; 10.
DR	SMART; SM00320; WD40; 10.
DR	PROSITE; PS00837; NACHT; 1.
DR	PROSITE; PS00676; WD_REPEATS_1; 10.
DR	PROSITE; PS00892; WD_REPEATS_2; 10.
DR	PROSITE; PS00294; WD_REPEATS_REGION; 1.
DR	Repeat; WD_repeat.
SQ	SEQUENCE 1356 AA; 149694 MW; 572A0B034DCE0D21 CRC64;

Query Match	16.3%; Score 442; DB 3; Length 1356;
Best local Similarity	27.6%; Pred.No. 1.7e-22;
Matches 130; Conservative	71; Mismatches 168; Indels 102; Gaps 15;

QY	56 GLQYVEA-----EVSINEDG-TLFDGRPIESLIDAVMPDVVQTQQAYRDKLAQQQA	108
DB	: :	
DB	875 GTQTLEGHGGSVMSVAFSPDGQRVASGSDDKTIKWDAASGTCTQT-----LEGHGS	926
QY	109 AAAAAAAAASQGSAXNGE-----NTANGEANGAHTIANNHYDMMEVDGDVEIPPNKA	162
DB	: :	
DB	927 SVLSVAFSPDGQRVASGSGDKTIKWDATAG--TGTQT-----	962
QY	163 VVLRGHESEVFICAMNPFVSDLLASGDSSTARWNLSENSTSGS-TOLVLRHCITREGQD	221
DB	: :	
DB	963 --LEGHGGSVMSVAFSPDGQRVASGSGDKTIKW----DTASGTCTQLGSH-----	1008
QY	222 VPSNKQVTSLDNSEGLLATGYDGFARIW-TKDGNLASTLGQHKGPIPALKNWNKKGNF	280
DB	: :	
DB	1009 ---GNSVMSVAFSPDGQRVASGSDDKTIKIWDATASGTCTQTLEGHGGVMSVAFSPDGQR	1065
QY	281 ILSAGVDKTTIIIDAHTGEAKQCFPHSAFALDVDQSNNTF-----ASCSTDMDCIHV	333
DB	: :	
DB	1066 VASGISDGTIKIWDAASGTCTQLEHG-----DWQSVAFSPDGQRVASGSDDHITI	1119
QY	334 CKLGQDRPIKFTOGHTNEVNNAIKWDP TGNLLASCSDMTLIKWSMKQDCNCVHDLOQHKE	393
DB	: :	
DB	1120 WDAASGTCTQLEHGDSVMSVAFSPDGQRVASGSDGTIKIWDAASGTCTQLEGHGGW	1179
QY	394 IYTIKNSPTPGTNPNANMLASAFSDSVIRLWDVDRGICITLKHOEPVYSVAFSPD	453
DB	: :	
DB	1180 VHSVAFSPDQG-----RVASGSDIGTIKWDAASGTCTQLEGHGGVMSVAFSPD	1230
QY	454 GRVLASGSPKCVHIW-----NTQVCL--HYLNQVQLINLGR	488
DB	: :	
DB	1231 QORVASGSSDNTIKIWDATASGTCTQLNVGSTATCLSPDYTNAYINTNIGR	1281

RESULT 15	
Q8X1P5	
ID	PRELIMINARY; PRT; 1356 AA.
AC	O8X1P5;
DT	01-MAR-2002 (TrEMBLrel. 20, Created)
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Beta transducin-like protein HET-E2C.
GN	HET-E.
OS	Podospira anserina.
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC	Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX	NCBI_TaxID=5145;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Espagne E., Balhadere P., Begueret J., Turcq B.;
RT	"A new family of WD40 proteins implicated in vegetative
RT	incompatibility; evidence for a major role of WD40 repeat domain in
RT	the specificity of het-e and het-d genes";
RL	Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AF323582; AAL37298.1; -
DR	InterPro; IPR007111; NACHT_NTPase.
DR	InterPro; IPR001680; WD40.
DR	Pfam; PF05729; NACHT; 1.
DR	Pfam; PF00400; WD40; 10.

DR PRINTS; PRO0320; GPROTEINERPT.
DR PRODom; PD000018; WD40; 10.
DR SMART; SMC0320; WD40; 10.
DR PROSITE; PS00837; NACHT; 1.
DR PROSITE; PS00678; WD REPEATS_1; 10.
DR PROSITE; PS00082; WD REPEATS_2; 10.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 1356 AA; 149720 MW; 9A80777304B361D0 CRC64;

Query Match 16.3%; Score 442; DB 3; Length 1356;
Best Local Similarity 27.6%; Pred. No. 1.7e-22;
Matches 130; Conservative 71; Mismatches 168; Indels 102; Gaps 15;

QY 56 GLOVVEA-----EVSINEDG-TLFDGRPIESLSLDVMPDVVQTRQQAAYRDKLAQQA 108
DB 875 GTQLEHGGSVWVAFSPDQGVASGSDDKTIKIWDASGCTCTQ-----LEGHGS 926
QY 109 AAAAAAQAQGSARNGE-----NTANGENGAAHTIANHNTMMVGDVGEIIPPNA 162
DB 927 SVLSVAFSPDQGVASGSDKTIKINDTAGS--TGTQ----- 962
QY 163 VVLRCHSEVPICAWNPVSDLLASGSDSTARINLSENSTSGS-TQLVLRHCIREGGQD 221
DB 963 --LEGHGSVWVAFSPDQGVASGSDKTIKIW----DPSGCTCTQLEGH----- 1008
QY 222 VPSNKDVTSLDWNSSEGTLLATGSYDGFARIW-TKDGNLASTLGHKGPFPALKWKKGNF 280
DB 1009 ---CNSVWVAFSPDQGVASGSDDKTIKIWDIASGCTCTQLEHGGSVWVAFSPDQGR 1065
QY 281 ILSAGVDKTIWDIAHTGEAKQPPPHSAPALVDWQSNNTF-----ASCSTDICIHV 333
DB 1066 VASGSDGTIKINDAASGCTCTQLEHG-----DWQSVAFSPDQGVASGSDDHITIKI 1119
QY 334 CKLGQDRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQOHKE 393
DB 1120 WDAASGCTCTQLEHGGSVWVAFSPDQGVASGSDGTIKIWDASGCTCTQLEHGGSW 1179
QY 394 IYTIKWSPTGPTNPNANMLASAPDSTVRLWVDVDRGICIHITLKHQEPVYVAFSPD 453
DB 1180 VHSVAFSPDQ-----RVASGSDGTIKIWDASGCTCTQLEHGGSVWVAFSPD 1230
QY 454 GRYLASGSFDCVHIW-----NTQVCL--HYLNGQVLLNLGR 488
DB 1231 GQVASSGSDNTIKINDIASGCTCTQTLNVGSTATCLSFDTNAYINTNIGR 1281

Search completed: August 9, 2004, 16:48:10
Job time : 47.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:41:43 ; Search time 13.3333 Seconds
(without alignments)
3708.183 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 2731

Sequence: 1 MSISSEVNFVLYRQESG.....GKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501.5	18.4	535	2	S48268
2	481	17.6	1526	2	AC2239
3	462	16.9	1356	2	T18521
4	459.5	16.8	1258	2	AI2155
5	451	16.5	1683	2	AF2071
6	437	16.0	1227	2	AF1810
7	436.5	16.0	1189	2	AI2493
8	436.5	16.0	1708	2	AE1866
9	415	15.2	1711	2	AD1842
10	404.5	14.8	1747	2	AC1842
11	403.5	14.8	934	2	AG1889
12	391.5	14.3	564	2	T40883
13	390.5	14.3	1693	2	S76086
14	385.5	14.1	676	2	AE2195
15	385	14.1	1551	2	AB2410
16	374.5	13.7	515	2	S19487
17	373	13.7	1189	2	AH2154
18	369.5	13.5	502	2	T41148
19	366.5	13.4	304	2	AG1837
20	362	13.3	265	2	AF1890
21	358	13.1	437	2	S05357
22	352.5	12.9	677	2	AE1861
23	344.5	12.6	786	2	AG2375
24	340	12.4	333	2	G85034
25	339.5	12.4	559	2	AB2202
26	334.5	12.2	410	2	S48052
27	331	12.1	323	2	T02617
28	331	12.1	777	2	T41075
29	329.5	12.1	409	2	S36113

ALIGNMENTS

RESULT 1

S48268 probable membrane protein YBR103w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBR0832

C:Species: Saccharomyces cerevisiae

C>Date: 01-Aug-1995 #sequence revision 01-Sep-1995 #text_change 19-Apr-2002

C:Accession: S48268; S45971; S44683

R:Mannhaupt, G.; Stucka, R.; Ehnlé, S.; Vetter, I.; Feldmann, H.

Yeast 10, 1363-1381, 1994

A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.

A:Reference number: S48255; MUID:95208357; PMID:7900426

A:Accession: S48268

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-535 <MAN>

A:Cross-references: EMBL:X78993; NID:g476045; PIDN:CAAS5606.1; PID:g476059

R:Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45927

A:Accession: S45971

A:Molecule type: DNA

A:Residues: 1-535 <PE2>

A:Cross-references: EMBL:Z35972; MIPS:YBR103w

C:Genetics:

A:Gene: SGD:SIF2

A:Cross-references: SGD:S0000307

A:Map position: 2R

C:Superfamily: unassigned WD repeat proteins; WD repeat homology

C:Keywords: transmembrane protein

F:355-388/Domain: WD repeat homology <WD1>

F:397-429/Domain: WD repeat homology <WD2>

F:434-450/Domain: transmembrane #status predicted <TMM>

Query Match 18.4%; Score 501.5; DB 2; Length 535;

Best Local Similarity 25.4%; Pred. No. 1.2e-28;

Matches 143; Conservative 101; Mismatches 239; Indels 79; Gaps 16;

Qy	1	MSISSEVNFVLYRQESGFSHSAFTFGTSHISQSNIN-GALVPPAALISIIQKGLQY	59
Db	1	MSITSEELNYLWKYQCEMGHEVSALALQDETREVLEFDEKYEHIPLGLTVNLVQRGILY	60
Qy	60	VEAEVSIQEDG---TLFDGRPIESLSLIIDAVMPDVVQTRQQAYRDKLAQHAAAAAAA	116
Db	61	TESELMVDSKGDTSALNEHHLSDFNLVQALQID-----KERFPE--ISSEGRFTL	109
Qy	117	ATTQQCSAKNGENTANGEANGHTIANNHTDMN-EVDGVEIFSNKAVLVRGHESSEVFC	175
Db	110	ETNSE-SNKAGEDGASTVERETQEDDTNSIDSSDDLGGFKL--LKEIV---KLDNIVSS	163
Qy	176	ANPVPD-LIVSGSGDSTARIWNLSNENSGTPQ-----LVLRHCIRGGQVPSNKDV	228
Db	164	TWNPLDESILAYGKNSVARLARIVETDQEGKKYKWLITIAELRHPPFALSASSGKTTNQV	223

30	328.5	12.0	606	2	T08180
31	328	12.0	589	2	AG2400
32	327.5	12.0	473	2	T33805
33	326	11.9	798	2	S34023
34	325.5	11.9	787	2	T00798
35	324.5	11.9	579	2	T22703
36	324	11.9	342	2	AE2490
37	313	11.5	640	2	S49932
38	312	11.4	357	2	AI2099
39	312	11.4	376	2	T19266
40	312	11.4	1049	2	T42045
41	311.5	11.4	317	2	T46032
42	310	11.4	1191	2	S76414
43	307.5	11.3	704	2	S33263
44	307	11.2	643	2	T41454
45	303.5	11.1	501	2	T27513

PF20 protein, micr
WD-repeat protei
hypothetical prote
TATA box-binding p
hypothetical prote
hypothetical prote
WD-repeat protei
MET30 protein - ye
WD-40 repeat prote
hypothetical prote
beta transducin-li
WD-40 repeat regul
beta transducin-li
transcription init
transcription init
hypothetical prote

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2071
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1258
A:Cross-references: GB:BA000019; PIDN:BA074499.1; PID:gl7131893; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2800

Query Match 16.8%; Score 459.5; DB 2; Length 1258;
Best Local Similarity 31.5%; Pred. No. 4.6e-25;
Matches 102; Conservative 44; Mismatches 101; Indels 77; Gaps 7;
QY 165 LRGESEVFICAMNPVSDLLVSGSGSTARIWNLSENSTSGTQVLVLRHCIRREGQDVPS 224
DB 974 LEGHTDFIYGLAFSPDSQTLASASTSSVRLNMI---STGQCQFILLER----- 1019
QY 225 NKDVTSLDW-----NSEGTLLATGSDYGFARIWT-KDGNLASTLQGHKGPFLAKWKKK 277
DB 1020 -----TDVYAVVPHQKIATGSDCTVKLWNLSTGQCLKTLSEHSDKILGNWSPD 1073
QY 278 GNFILSAGVDKTTIWDHATGKAKQFPFHSAPALDVMQSNNTFASCTDMCIHVCKLG 337
DB 1074 GOLLASASADQSVRLWD-----CCTGRCVGI----- 1099
QY 338 QDRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKEIYTI 397
DB 1100 -----LRGHNRYVSAIFSPNGEIIATCSDQIVKIWDWQGGCKLTGHTNWVDDI 1152
QY 398 KWSPTGPTNNPNANMLASASDFSVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYL 457
DB 1153 AFSPDGK-----ILASASHDQVRIWDVNTGKCHICIGHTHLVSSVAFSPDGEV 1203
QY 458 ASGSPDKCVHINVTQALVHSYR 481
DB 1204 ASGSQDTVRIWNVTGECQLIIR 1227

RESULT 5
AF2071
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AF2071
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2071
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1683
A:Cross-references: GB:BA000019; PIDN:BA073823.1; PID:gl7131215; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2124

Query Match 16.5%; Score 451; DB 2; Length 1683;
Best Local Similarity 28.5%; Pred. No. 2.9e-24;
Matches 109; Conservative 71; Mismatches 163; Indels 40; Gaps 7;
QY 124 AKNGENTANGENGHAHTIANNHTDMMEVDGVEIPSNKAVLVRGHESEVFICAMNPVSD 183
DB 1290 SSDGKAIASASRDNTIKLWNRHGLE-----TFTGSHGGVAVVNFPLPDSNI 1336
QY 184 LVSGSGDSTARIWNLSENSTSGTQVLVLRHCIRREGQDVPSNKDVTSLDWNSEGTLLATG 243
DB 1337 TASAIDNTIRLW---QRLISPLEV-----LAGNSGVYAVSFLHDGSIATA 1381

QY 244 SYDGFARIW-TKDGNLASTLQGHKGPFLAKWKNKGNFILSAGVDKTTIWDHATGSAK 302
DB 1382 GADGNIQLWHSQDGSLLKTLFNGKA-IYGISFTQGDLLIASANADKTVKIWRVRDGRALK 1440
QY 303 QPPFHSAPALDVMQ-SNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPTG 361
DB 1441 TLIGHDNEVKNVSPDGKTLASASRDNTVKLWNVSDGKFKKTLKGHTDEVPVYFSPDG 1500
QY 362 NLLASCSDDMTLKIWSMKQDNCVHDLOAHNKEIYTIKWSPTGPTNNPNANMLASASFD 421
DB 1501 KIIASASADKTIIRLWDSPGNLKLKSLFAHNDLVSVNFPDGS-----MLASTSAD 1551
QY 422 STVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYLASGFDKCVHINVTQGTALVHSYR 481
DB 1552 KTVKLRSHDHGHLHTSGHNVVYSSFSFDGRIYIASASEDKTVKIWDHGLTTLTPQ 1611
QY 482 GTGGEFVYVNAAGDKVGSASD 504
DB 1612 HQAGVMSAIFSPDGKTLISGLSD 1634

RESULT 6
AB1810
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB1810
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB1810
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1227
A:Cross-references: GB:BA000019; PIDN:BA077553.1; PID:gl71315007; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all+0029

Query Match 16.0%; Score 437; DB 2; Length 1227;
Best Local Similarity 31.7%; Pred. No. 1.9e-23;
Matches 102; Conservative 61; Mismatches 129; Indels 30; Gaps 7;
QY 157 IPSNKAVLVRGHESEVFICAMNPVSDLLVSGSGSTARIWNLSENSTSGTQVLVLRHCIR 216
DB 888 ITGNSMLILRGYTRDVYVAFSPDSQTLASGRDDYITGLWNLKTGEC-----HPLR 938
QY 217 EGGQDVPSNKDVTSLDWNSEGTLLATGSDYGFARIW---TKDGNLASTLQGHKGPFL 272
DB 939 -GHQ-----RIRSVAFHPDGKILASGSADNTIKLWDISDTNHSKYIRTLTGHTNWVTV 992
QY 273 KWNKGNFILSAGVDKTTIWDHATGKAKQFPFHSAPALDVMQ-SNNTFASCTDMCI 331
DB 993 VFSPDKHTLASSSEDRIRLWDKDTGCLQKLGHSWWVTVAFSPDGRILASGSADSEI 1052
QY 332 HVCKLGQDRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOAHN 391
DB 1053 KIMDVASGKCLQTLTDPQGMVSVAFSLDGTLLASASEDQTVKLNKLTGECVHTLKGE 1112
QY 392 KEIYTIKWSPTGPTNNPNANMLASASDFSVRLWDVDRGICHTLTKHQEPVYVAF 450
DB 1113 KQVYVAFSPNGQ-----IAAGSEDTTVKLMWDISTGSCVDTLKHGHTAAIRSVAF 1163
QY 451 SPDGRYLASGFDKCVHINVTQ 472
DB 1164 SPDGRLLASGSEDEKTLQWDMQ 1185

RESULT 7
AI2493

WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AF1493
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AF1493
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1189 <KUR>
 A:Cross-references: GB:BA000020; PIDN:BA78213.1; PID:gi71135667; GSPDB:GN00180
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr7129
 A:Genome: plasmid

Query Match 16.0%; Score 436.5; DB 2; Length 1189;
 Best Local Similarity 28.7%; Pred. No. 2e-23;
 Matches 108; Conservative 65; Mismatches 142; Indels 61; Gaps 9;

QY 167 GHESEVFCANVPVSDLLVSGSDSTARIWNLSE-----NSTSGPTQLVLRHCIRGGQ-- 220
 DB 686 GHDACVMSVVFHPVGQILATAGEDNTIKLWELQSGCCLTLQGHQHWVWTFIAFNSGGRIL 745
 QY 221 -----DVPNSK-----DVTSIDMNSGTLTATGSDYDGFARIW-TKQGN 257
 DB 746 ASGSPQNVKLVDIHTGCKVMTLQGHGTGVTVAFPKDNLLSGSDQSVKVMKRTGR 805
 QY 258 LASTLQGHKGPFPALKKWKNGFILSAGVDKTTIWDANTGEAKQFPFHSAPALDV--D 315
 DB 806 CLDTLKKHTNRIWSVAFHPQGLFVSGDDHAAKIWEIQTGQCIKTFQGHSNATYIAHN 865
 QY 316 WQSNNTFASCTDWCIVHCKLG-----QDRPKTFQGHNEVNAIKWDPGTGNLLAS 366
 DB 866 WE-HSLLASGHEDQTKLWDLNLSHPKSNVNTHPRIILQHSNRVSVFVSTGQLLAS 924
 QY 367 CSDMTLKIWSKQDNKVDHDLQAHNKEIYTIKSPGTGPTGNPNPNAIMLASFSTVRL 426
 DB 925 GSADRTIKLWSPHTGQCLTLHGSGWMAIAFSLD-----DKLIASGSYDHTVKI 975
 QY 427 WDVDRGTCIHTLKHQEPVVSVAFPDGRVLAGSPDKCVHIWNTOTGALVHSYRGTGI 486
 DB 976 WDVSSGQCLTLQGHFGSVLAVAFSCDKLTFSSGKELVKQWDETGYCLQT----- 1028
 QY 487 FEVCWNAAGDKVGASA 502
 DB 1029 -----WEADSNRVNVA 1040

RESULT 8
 AE1866
 WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AF1866
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AF1866
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1708 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA72436.1; PID:gi7129823; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all0478

Query Match 16.0%; Score 436.5; DB 2; Length 1708;
 Best Local Similarity 30.5%; Pred. No. 3.3e-23;
 Matches 123; Conservative 70; Mismatches 169; Indels 41; Gaps 15;

QY 119 NOGSANNGENTANGBENGAAHTIA--NNHT-DMMEVDGDEIPEPSKAVV--LRGHESE 171
 DB 1298 NRQG---NLKTLISGHTAGVATVTFPNGETIGSASIDATLKLWSPQGLLGLTKGHSW 1354
 QY 172 VFICAMNPVSDLLVSGSDSTARIWNLSENSTSGPTQLVLRHCIRREGQDVPSNKDVTSL 231
 DB 1355 VNSVSPSPDGRIFASGSRDKTIVTLRWDE-----VLLRNPKGDG-----NDWVTSI 1400
 QY 232 DWSSEGTLATGSDYDGFARITWKDGNLSTLQGHKGPFPALKKWKNGNFILSAGVDKTTI 291
 DB 1401 SFSDDGTTLAASRDQTVKILSRHGLKLLTFKGTGSIWGVAMSPNRQWIASAKDQTVK 1460
 QY 292 IWDATGEAKQFPFHSAPALDVWQSN--TFASCSTDMCIHVCKLGQDRPKITFGHTN 350
 DB 1461 LMW-QDGKILHTLQGHODAVLAVANSSDQSVIASAGDKIVKIWSQG-QGLLHTLQGH 1518
 QY 351 EVNAIKWDPGTGNLLASCDSDMTLKIWSKQDNKVDHDLQAHNKEIYTIKSPGTGPTGN 410
 DB 1519 AVNWVSFSPDGKILASVSDDTTVKLWS-RDQQLLHTLKEHSRRVNGVAMSPDQ-- 1571
 QY 411 ANMLASASPDSTVRLWVDVRGICHTLTKHQBPTVSVAFPDGRVLAGSPDKCVHIW 470
 DB 1572 ---ILASASIDGTVKLNWRD-GSLSRNLPGDGSFISVSPDGKMLAANSDDQ-IRLMN 1626
 QY 471 TQTGALVHSYRG-TGGIFFCVWNAAGDKVGASASDQSVCLDL 512
 DB 1627 -QKGTLLMVLKGDDELTSVTFSPDSQILAVGGNGKVFIFNL 1668

RESULT 9
 AD1842
 WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AD1842
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD1842
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1711 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA77808.1; PID:gi7135262; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all0284

Query Match 15.2%; Score 415; DB 2; Length 1711;
 Best Local Similarity 24.9%; Pred. No. 1.2e-21;
 Matches 123; Conservative 75; Mismatches 180; Indels 116; Gaps 14;

QY 65 SINEDGTL----PDRPIESLILDAVMPDVVQTRQAYRDKLAQQAHAHAATAATNQ 120
 DB 1246 SASDDGITRLWSLDRPLITI-----PSHTKQVLAVTFSPDG 1282
 QY 121 QGSANNGENTANGBENGAAHTIANNHTDMMEVDGDEIPEPSKAVVLRGHESEVFCANVP 180
 DB 1283 QTIVSAG-----ADNTVKLWSRNGTL-----LTTLEGHNEAVQVIFSPD 1322
 QY 181 SDLLVSGSDSTARIWNLSENSTSGPTQLVLRHCIRREGQDVPSNKDVTSLDWNSEGL 240
 DB 1323 GRLTASADKTIILSRDGN-----IL-----GTFAGHNEVNSLSFSPDGNIL 1367
 QY 241 ATGSYDGFARITWKDGNLSTLQGHKGPFPALKKWKNGNFILSAGVDKTTIINDA----- 295

```

Db 1368 ASGSDNTVRLWTVNRTLPKTFYGHKGSVSYVRFNSDNGKKTISLSTDSMTKWSLDGKLL 1427
QY 296 -----HT-----GEAKQQFFPHGAPALDWDQSN- 319
Db 1428 QTLSSPLPDVTSISFTPDNKIVALASPDHTIHLNRYQGGLRLSLPCHNHWITSLSFSPNK 1487
QY 320 NTFASCSTDMCHVCKLQODRIKTPQGHNEVNAIKWDPTGNLLASCDDMTLKISWMK 379
Db 1488 QILASGSADTKIKLSV-NGRLLKTLKGHWVTDIKFSADGKNIVSASADTKIKLSL- 1545
QY 380 QNCVHDLQAHNKEIYTIKWSPTGPTNNPNANMLASASFDSTVRLWDVDRGICIHILT 439
Db 1546 DGLRTLTLOHGSASVSNLSPDQ-----TLASTQDETILKNLNL-GELIYILR 1595
QY 440 KIQEPYVSFAFDGRYLAGSFDKCVHIWNTQTGALVHSYRG-TGGIFEVCWNAAGDKV 498
Db 1596 GHSDDVYNLSFSPDGTIASASDDGTIKLWNPENGTLTKTQCHRGVRSVSFSPDGKIL 1655
QY 499 GASASDGSVCVLDL 512
Db 1656 ASGGHDTTVKWNL 1669

RESULT 10
AC1842
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
R:Accession: AC1842
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1747 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077807.1; PID:gl7135261; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0283

Query Match 14.8%; Score 404.5; DB 2; Length 1747;
Best Local Similarity 26.9%; Pred. No. 7.2e-21;
Matches 112; Conservative 67; Mismatches 158; Indels 79; Gaps 10;
QY 165 LRHGESEVFCAMNPVSDLLVSGSGDSTARIWNL-----SENSTSGPTQ 208
Db 1310 LTGHRKRTSVKFSFSPDGKILASASGDKTIKFNTDGFELKTIAAHNQOVNSINFSSDSKT 1369
QY 209 LV-----LRHCIREGGQDV-----PSNK----- 226
Db 1370 LVASGADSTMKWKIDGLTIKITSRGQIRDVTFSPDNKVIASASSDKTVRIQLNQK 1429
QY 227 -----DVTSLDNSEGLTATSGYDFARITWKGDLA-----STLQCHKGPFAKWNKKG 278
Db 1430 SKSNVNSVSNFNPDKTFASAGWGNITIQPRE-TLAHSSLSTIQKNQNIITVSYSYSPDG 1488
QY 279 NFILSAGVDKTIINDAHTGEAKQFPFHSAPALDWDQ-SNNTFASCSTDMCHVCKLQ 337
Db 1489 KTIATASADNTIKLWDSQTQQLIKTLTGHKORITTLTFHPNQTIASGSADTKIKWRN 1548
QY 338 QDRPIKTPQGHNEVNAIKWDPTGNLLASCDDMTLKISWMKQNCVHDLQAHNKEIYTI 397
Db 1549 DQQLRTLTGHNDVTSVNFSPDGGFLASGSDTNVTKW-QTDGELIKNIITGHGLATSV 1607
QY 398 KWSPTGPTNNPNANMLASASFDSTVRLWDVDRGICIHILTQKQEPYVSFAFDGRYL 457
Db 1608 KFSFD-----SHTILASASWNTIKLWQVTDGKILNINLNGHIDGVITLSLSFSPDGEIL 1658
QY 458 ASGSFDKCVHIWNTQTGALVHSYRG-TGGIFEVCWNAAGDKV GASASDGSVCVLDL 512

```

```

Db 1659 ASGSADNTIKLWNLPNATLLKTLGHGKINTLAFSPDGKTLTSGGEDAGVWVWNL 1714

RESULT 11
AG1889
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
R:Accession: AG1889
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1889
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-934 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072622.1; PID:gl7130010; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0664

Query Match 14.8%; Score 403.5; DB 2; Length 934;
Best Local Similarity 23.1%; Pred. No. 3.5e-21;
Matches 128; Conservative 115; Mismatches 197; Indels 113; Gaps 21;
QY 8 VNFVLYRYLQESGFSHSAFTFGTIESHSQSNINGALVPPAALISIIQKQYVEAEVYSIN 67
Db 231 ISSLVTLGIIALGLAGVANTQQKARNSETK-----AISSAESFLNANLEF- 277
QY 68 EDGLTFD---GRITESLSDA-VMPDVVQTRQAY-----RDKLAQCHAAAAAATAAAT 118
Db 278 -DGLIASIRAGRRIKGTGDIDANTRTQITETLQOS-NFVREKRLAEHDGMLSEVSFSPD 336
QY 119 NQGSASAKNGENTANGEGAHTIANNHTDMVEVDGDIPEISNKAVLVRGHSEVF-ICAW 177
Db 337 SKFIATASRQTKVK-----IWSLDGKKQL-----VLRBEKGGEGFNSVAF 376
QY 178 NPVSDLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGO--DVPSNKD-VTSLDN 234
Db 377 SPDGTLMATGSDWNTAKINS-----REGKRLHTLDGKHEAVLEVAFS 418
QY 235 SEGTLIATSGYDFARITWKGDLASTLGOHKGPFAKWNKGNFILSAGVDKTIWID 294
Db 419 PDSQLIATASWNTVKLWSREGKLLHTLEGHKDKVNSITFSPDGQIATVGVWNTMKLWN 478
QY 295 AHTGEAKQFPFHSAPALDWDQ-----SNNTFASCSTDMCHVCKLQODRIKTPQGH 349
Db 479 L-DGKELRTFRGHQ----DMIWSVSFSPDGKQIATASGDTVKLWSL-DGKELQTLRGHQ 532
QY 350 NEVNAIKWDPTGNLLASCDDMTLKISWMKQNCVHDLQAHNKEIYTIKWSPTGPG---- 405
Db 533 NGVNSVTFSPDGKLIATASGDTVKLWNSKGQF-LETLYGHTDAVNSVAFSPDGTSIATA 591
QY 406 -----TNNPNANLM-----LASASFDSTVRLWDV--DRGICI 435
Db 592 GNDKTAIKWLNKSPNSIIVRGHEDEVDFDLVFSPNGKYIATASWDTAKLMSIVGDKLQEL 651
QY 436 HTLTKQEPYVSFAFDGRYLAGSFDKCVHIWNTQTGALVHSYRG-TGGIFEVCWNA 494
Db 652 RTFNGHQVRNKLFSFSPDGKYIATTSWDTAKLWNLD-GTLQKTLTGKDKTVNSVNFSPD 710
QY 495 GDKVGASASDGSV 507
Db 711 GLIATASEDKTV 723

RESULT 12
T40883
WD repeat protein - fission yeast (Schizosaccharomyces pombe)

```

RESULT 13
S76086
beta transducin-like protein, 190K - *Synechocystis* sp. (strain PCC 6803)
N:Alternate names: protein sl0163
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 26-May-2000
C:Accession: S76086
R:Kaneko, T., Sato, C., Hasegawa, S., Kotani, S.; Kotani, S.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res 3, 109-136, 1998



GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:41:13 ; Search time 34.6667 Seconds
(without alignments)
4678.164 Million cell updates/sec

Title: US-09-987-701-12
Perfect score: 2731
Sequence: 1 MSISSEVFLVRYLQESG.....GDKVGASASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2731	100.0	514	11 Q9EQD4	Q9eqd4 mus musculus
2	2727	99.9	514	11 Q8BHJ5	Q8bhj5 mus musculus
3	2724	99.7	514	11 Q8CBG4	Q8cbg4 mus musculus
4	2645.5	96.9	519	13 Q7SZM9	Q7szm9 xenopus lae
5	2466.5	90.3	527	11 Q8BYQ4	Q8byq4 mus musculus
6	2462.5	90.2	527	11 Q8BMM0	Q8bmm0 mus musculus
7	2439	89.3	577	4 Q8EUY2	Q8euy2 mus musculus
8	1956	71.6	412	11 Q8C0A1	Q8c0a1 mus musculus
9	1906.5	69.8	700	5 Q9XZK1	Q9xzki drosophila
10	1388	50.8	613	10 Q9FN19	Q9fnl9 arabidopsis
11	1101	40.3	201	11 Q8VEG3	Q8veg3 mus musculus
12	920.5	33.7	524	5 Q8ERJ9	Q8erj9 drosophila
13	480	17.6	1356	3 Q8X1P4	Q8x1p4 podospora a
14	477	17.5	1356	3 Q8X1P5	Q8x1p5 podospora a
15	472	17.3	1356	3 Q8X1P3	Q8x1p3 podospora a
16	445	16.3	1376	3 Q8X1P2	Q8x1p2 podospora a

17	437	16.0	1227	16 Q8Z0R1	Q8z0r1 anabaena sp
18	436.5	16.0	1189	16 Q8YL09	Q8yl09 anabaena sp
19	436.5	16.0	1708	16 Q8YZI2	Q8yzi2 anabaena sp
20	415	15.2	1711	16 Q8Z019	Q8z019 anabaena sp
21	404.5	14.8	1747	16 Q8Z020	Q8z020 anabaena sp
22	403.5	14.8	934	16 Q8YZ23	Q8yz23 anabaena sp
23	391.5	14.3	564	3 Q74845	Q74845 schizosacch
24	385.5	14.1	676	16 Q8YSG6	Q8ysg6 anabaena sp
25	385	14.1	1551	16 Q8YMU3	Q8ymu3 anabaena sp
26	376	13.8	415	4 Q8N136	Q8n136 homo sapien
27	373	13.7	1189	16 Q8YTD1	Q8ytd1 anabaena sp
28	370.5	13.6	1430	16 Q98HK1	Q98hk1 rhizobium l
29	370	13.5	1233	17 Q8TMY4	Q8tmx4 methanosarc
30	369.5	13.5	502	3 Q74855	Q74855 schizosacch
31	368	13.5	415	4 Q8N776	Q8n776 homo sapien
32	366.5	13.4	304	11 Q9D4T2	Q9d4t2 mus musculus
33	366.5	13.4	304	16 Q8Z054	Q8z054 anabaena sp
34	366.5	13.4	481	5 Q9VPR4	Q9vp4 drosophila
35	366.5	13.4	488	5 Q8T4A2	Q8t4a2 drosophila
36	365	13.4	411	5 Q96698	Q96698 drosophila
37	362	13.3	265	16 Q8YZ16	Q8yz16 anabaena sp
38	361.5	13.2	1241	2 Q9XBD8	Q9xbd8 amycolatops
39	358.5	13.1	480	5 Q96995	Q96995 drosophila
40	358	13.1	478	5 Q861S4	Q861s4 dictyosteli
41	357	13.1	476	13 Q93531	Q93531 xenopus lae
42	357	13.1	476	13 Q7ZXK9	Q7zxk9 xenopus lae
43	355	13.0	339	11 Q9DCZ7	Q9dcz7 mus musculus
44	355	13.0	339	11 Q99KN2	Q99kn2 mus musculus
45	355	13.0	352	10 Q80990	Q80990 arabidopsis

ALIGNMENTS

RESULT 1

ID	Q9EQD4	PRELIMINARY;	PRT;	514 AA.
AC	Q9EQD4;			
DT	01-MAR-2001 (Tremblrel. 15, Created)			
DT	01-MAR-2001 (Tremblrel. 15, Last sequence update)			
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)			
DE	IRAL.			
GN	IRAL OR 803049H02RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/c;			
RA	Zhang X., Dormady S., Basch R.;			
RT	"Identification of four human cDNAs that are differentially expressed by early hematopoietic progenitors.";			
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF268195; AAC44738.1; -			
DR	MGD; MGI:2441730; Iral.			
DR	InterPro; IPR006594; Lish.			
DR	InterPro; IPR001680; WD40.			
DR	Pfam; PF00400; WD40; 8.			
DR	PRINTS; PR00320; GPROTEINRPT.			
DR	ProDom; PD000018; WD40; 3.			
DR	SMART; SM00567; Lish; 1.			
DR	SMART; SM00320; WD40; 8.			
DR	PROSITE; PSS0896; Lish; 1.			
DR	PROSITE; PSS0678; WD_REPEATS_1; 4.			
DR	PROSITE; PSS0082; WD_REPEATS_2; 6.			
DR	PROSITE; PSS0294; WD_REPEATS_REGION; 1.			
KW	Repeat; WD repeat.			
SQ	SEQUENCE 514 AA; 55689 MW; 6A7CE68A40C141F CRC64;			

Query Match 100.0%; Score 2731; DB 11; Length 514;
Best Local Similarity 100.0%; Pred. No. 1.7e-184;
Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSISSDEVNVLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQV 60
 DB 1 MSISSDEVNVLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSNEDGTLDPGRPIESLSLIDAMPDVVQTRQAVRDKLAQCHAAAAAATNQ 120
 DB 61 EAEVSNEDGTLDPGRPIESLSLIDAMPDVVQTRQAVRDKLAQCHAAAAAATNQ 120
 QY 121 QGSAKNGENTANGEGAHTIANNHTDMVEVDGVEIPSNKAVLRGHESEVEFICAWNPV 180
 DB 121 QGSAKNGENTANGEGAHTIANNHTDMVEVDGVEIPSNKAVLRGHESEVEFICAWNPV 180
 QY 181 SLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 DB 181 SLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 QY 241 ATGSYDGFARITWKDGNLSTLQGHKGPFIKWKKNKGNFIILSAGVDKTTIWDHTGEA 300
 DB 241 ATGSYDGFARITWKDGNLSTLQGHKGPFIKWKKNKGNFIILSAGVDKTTIWDHTGEA 300
 QY 301 KQOPPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPKTFQCHTNEVNAIKWDPT 360
 DB 301 KQOPPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPKTFQCHTNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASAF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASAF 420
 QY 421 DSTVRLWDVDRGICHTLTKHQBPVSVAFSPDGRYLASGSDKCVHIWNTOTGALVHSY 480
 DB 421 DSTVRLWDVDRGICHTLTKHQBPVSVAFSPDGRYLASGSDKCVHIWNTOTGALVHSY 480
 QY 481 RGTGGIFEVCWNAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RGTGGIFEVCWNAAGDKVGASASDGSVCVLDLRK 514

RESULT 2

Q8BHJ5 Q8BHJ5 PRELIMINARY; PRT; 514 AA.
 ID Q8BHJ5
 AC Q8BHJ5
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE IRAL protein.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RA "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK029595; BAC26526.1; -
 DR EMBL; AK033347; BAC28241.1; -
 DR PIR; PT0651; PT0651.
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR Pfam; PF00400; WD40; 8.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PSS0896; Lish; 1.
 DR PROSITE; PSS0678; WD_REPEATS_1; 4.
 DR PROSITE; PSS0678; WD_REPEATS_1; 4.

DR PROSITE; PSS0082; WD_REPEATS_2; 6.
 DR PROSITE; PSS0294; WD_REPEATS_REGION; 1.
 SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F8BF14 CRC64;
 Query Match 99.9%; Score 2727; DB 11; Length 514;
 Best Local Similarity 99.8%; Pred No 3.2e-184; Indels 0; Gaps 0;
 Matches 513; Conservative 0; Mismatches 1;
 QY 1 MSISSDEVNVLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQV 60
 DB 1 MSISSDEVNVLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSNEDGTLDPGRPIESLSLIDAMPDVVQTRQAVRDKLAQCHAAAAAATNQ 120
 DB 61 EAEVSNEDGTLDPGRPIESLSLIDAMPDVVQTRQAVRDKLAQCHAAAAAATNQ 120
 QY 121 QGSAKNGENTANGEGAHTIANNHTDMVEVDGVEIPSNKAVLRGHESEVEFICAWNPV 180
 DB 121 QGSAKNGENTANGEGAHTIANNHTDMVEVDGVEIPSNKAVLRGHESEVEFICAWNPV 180
 QY 181 SLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 DB 181 SLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 QY 241 ATGSYDGFARITWKDGNLSTLQGHKGPFIKWKKNKGNFIILSAGVDKTTIWDHTGEA 300
 DB 241 ATGSYDGFARITWKDGNLSTLQGHKGPFIKWKKNKGNFIILSAGVDKTTIWDHTGEA 300
 QY 301 KQOPPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPKTFQCHTNEVNAIKWDPT 360
 DB 301 KQOPPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPKTFQCHTNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASAF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASAF 420
 QY 421 DSTVRLWDVDRGICHTLTKHQBPVSVAFSPDGRYLASGSDKCVHIWNTOTGALVHSY 480
 DB 421 DSTVRLWDVDRGICHTLTKHQBPVSVAFSPDGRYLASGSDKCVHIWNTOTGALVHSY 480
 QY 481 RGTGGIFEVCWNAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RGTGGIFEVCWNAAGDKVGASASDGSVCVLDLRK 514

RESULT 3

Q8CBG4 Q8CBG4 PRELIMINARY; PRT; 514 AA.
 ID Q8CBG4
 AC Q8CBG4
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE IRAL protein.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RA "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK036064; BAC29294.1; -
 DR PIR; PT0651; PT0651.
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PSS0896; Lish; 1.
 DR PROSITE; PSS0678; WD_REPEATS_1; 4.
 DR PROSITE; PSS0678; WD_REPEATS_1; 4.


```
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; LISH; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00836; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS0082; WD REPEATS 2; 6.
DR PROSITE; PS00294; WD REPEATS REGION; 1.
SQ SEQUENCE 514 AA; 55689 MW; 13BED3753A725029; CRC64;

Query Match
Best Local Similarity 99.6%; Score 2724; DB 11; Length 514;
Matches 512; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSISDEVNLFVRYLOESGFHSFTFGIESHISQSNINGALVPPAALISIIQGLQYV 60
Db 1 MSISDEVNLFVRYLOESGFHSFTFGIESHISQSNINGALVPPAALISIIQGLQYV 60
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQCHAAAAAATNQ 120
Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQCHAAAAAATNQ 120
QY 121 QSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIIPSNKAVVLRGHESEVFIC 180
Db 121 QSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIIPSNKAVVLRGHESEVFIC 180
QY 181 SLLVSGSDSTARIWNLSNSTSGPTQLVRHCIREGGQDVPSNKDVTSLDWNS 240
Db 181 SLLVSGSDSTARIWNLSNSTSGPTQLVRHCIREGGQDVPSNKDVTSLDWNS 240
QY 241 ATGSDVGFARIWTKDGNLASTLGQHKGPITALKNNKGNFILSAGVDKTTIWDHTGEA 300
Db 241 ATGSDVGFARIWTKDGNLASTLGQHKGPITALKNNKGNFILSAGVDKTTIWDHTGEA 300
QY 301 KQGFPPHSPALPDVQSNNTFASCSTDMCHVCKLGQDRPIKTFQGTNEVNAIKWPT 360
Db 301 KQGFPPHSPALPDVQSNNTFASCSTDMCHVCKLGQDRPIKTFQGTNEVNAIKWPT 360
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANMLAS 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANMLAS 420
QY 421 DSTVRLWVDVRCIGIHTLTKEQEPVYSAFSDGRLASGSDKCVHIWNTQTGALVHSY 480
Db 421 DSTVRLWVDVRCIGIHTLTKEQEPVYSAFSDGRLASGSDKCVHIWNTQTGALVHSY 480
QY 481 RGTGGIFECVWNAAGDKVGASDGSVCVLDLRK 514
Db 481 RGTGGIFECVWNAAGDKVGASDGSVCVLDLRK 514

RESULT 4
Q7SZM9
ID Q7SZM9 PRELIMINARY; PRT; 519 AA.
AC Q7SZM9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nuclear receptor co-repressor complex subunit TBLR1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
CX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12794076;
RA Tomita A., Buchholz D.R., Obata K., Shi Y.B.;
RT "Fusion Protein of Retinoic Acid Receptor (alpha) with Promyelocytic
RT Leukemia Protein or Promyelocytic Leukemia Zinc Finger Protein
RT Recruits N-CoR-TBLR1 Corepressor Complex to Repress Transcription in
RT Vivo."
RL J. Biol. Chem. 278:30788-30795 (2003).

DR EMBL; AY225088; AAP20646.1; -.
KW Receptor.
SQ SEQUENCE 519 AA; 56043 MW; 5E998EDC8C92296 CRC64;

Query Match
Best Local Similarity 96.3%; Score 2645.5; DB 13; Length 519;
Matches 500; Conservative 2; Mismatches 12; Indels 5; Gaps 1;

QY 1 MSISDEVNLFVRYLOESGFHSFTFGIESHISQSNINGALVPPAALISIIQGLQYV 60
Db 1 MSISDEVNLFVRYLOESGFHSFTFGIESHISQSNINGALVPPAALISIIQGLQYV 60
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQCHAAAAAATNQ 117
Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQCHAAAAAATNQ 117
QY 118 --TNQGSANKGENTANGEANGAHTIANNHTDMMEVDGVEIIPSNKAVVLRGHESEVFIC 175
Db 118 --TNQGSANKGENTANGEANGAHTIANNHTDMMEVDGVEIIPSNKAVVLRGHESEVFIC 175
QY 176 ANPVSLLVSGSDSTARIWNLSNSTSGPTQLVRHCIREGGQDVPSNKDVTSLDWNS 235
Db 176 ANPVSLLVSGSDSTARIWNLSNSTSGPTQLVRHCIREGGQDVPSNKDVTSLDWNS 235
QY 181 ANPVSLLVSGSDSTARIWNLSNSTSGPTQLVRHCIREGGQDVPSNKDVTSLDWNS 240
Db 181 ANPVSLLVSGSDSTARIWNLSNSTSGPTQLVRHCIREGGQDVPSNKDVTSLDWNS 240
QY 236 EGTLLATGSDVGFARIWTKDGNLASTLGQHKGPITALKNNKGNFILSAGVDKTTIWDHTGEA 295
Db 236 EGTLLATGSDVGFARIWTKDGNLASTLGQHKGPITALKNNKGNFILSAGVDKTTIWDHTGEA 295
QY 296 HTGEAKQGFPPHSPALPDVQSNNTFASCSTDMCHVCKLGQDRPIKTFQGTNEVNAI 355
Db 296 HTGEAKQGFPPHSPALPDVQSNNTFASCSTDMCHVCKLGQDRPIKTFQGTNEVNAI 355
QY 301 HTGEAKQGFPPHSPALPDVQSNNTFASCSTDMCHVCKLGQDRPIKTFQGTNEVNAI 360
Db 301 HTGEAKQGFPPHSPALPDVQSNNTFASCSTDMCHVCKLGQDRPIKTFQGTNEVNAI 360
QY 356 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANML 415
Db 356 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANML 415
QY 361 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANML 420
Db 361 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANML 420
QY 416 ASASFDTVRLWVDVRCIGIHTLTKEQEPVYSAFSDGRLASGSDKCVHIWNTQTG 475
Db 416 ASASFDTVRLWVDVRCIGIHTLTKEQEPVYSAFSDGRLASGSDKCVHIWNTQTG 475
QY 421 ASASFDTVRLWVDVRCIGIHTLTKEQEPVYSAFSDGRLASGSDKCVHIWNTQTG 480
Db 421 ASASFDTVRLWVDVRCIGIHTLTKEQEPVYSAFSDGRLASGSDKCVHIWNTQTG 480
QY 476 LVHSYRTGGIFECVWNAAGDKVGASDGSVCVLDLRK 514
Db 476 LVHSYRTGGIFECVWNAAGDKVGASDGSVCVLDLRK 514
QY 481 LVHSYRTGGIFECVWNAAGDKVGASDGSVCVLDLRK 519
Db 481 LVHSYRTGGIFECVWNAAGDKVGASDGSVCVLDLRK 519

RESULT 5
Q8BYQ4
ID Q8BYQ4 PRELIMINARY; PRT; 527 AA.
AC Q8BYQ4
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transducin (Tbllx protein).
GN TBLLX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Hypothalamus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RA 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
```

RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh L.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schermer J.A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK038674; BAC30092.1; -;
DR EMBL; BC043105; AAH43105.1; -;
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SMO0667; Lish; 1.
DR SMART; SMO0320; WD40; 8.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
SQ SEQUENCE 527 AA; 56802 MW; 146435A9C51DFD6A6 CRC64;

Query Match 90.3%; Score 24652.5; DB 11; Length 527;
Best Local Similarity 87.5%; Pred. No. 1.6e-165;
Matches 463; Conservative 27; Mismatches 23; Indels 15; Gaps 3;

QY 1 MSISDEVNPLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQV 60
DB 1 MSITSDEVNPLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQV 60

QY 61 EAEVSINEDGTLFDGRIPIESLSLIDAVPDVQVOTROQAYRDKLAQQ--HAAAAAATAAT 118
DB 61 EAEISINEDGTVFDCRIPIESLSLIDAVPDVQVOTROQAFREKLTQQOANAAAAATAAT 120

QY 119 -----NQGSAGKNGENTANGENGATIANNHDTDMVEVDGVEIIPSKAVLVR 166
DB 121 ATSTAATTPAAAQQPPKNGEATVNGEANGAHAI--NNHSPKMEIDGVDVIPPSTATVLR 179

QY 167 GHESEVFCIAWNPVSDLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPSNK 226
DB 180 GHESEVFCIAWNPVSDLLASGSGDSTARINWLNENSGSTQLVLRHCIREGGHDVPSNK 239

QY 227 DVTSLDWSNSETLLATGSDGFARITWKGKNGLASTLQGHKGPFIKWKNGKNGIILSAGV 286
DB 240 DVTSLDWSNSETLLATGSDGFARITWKGKNGLASTLQGHKGPFIKWKNGKNGIILSAGV 299

QY 287 DKTTIILDAHTGEAKQOFPFHSAPALDVDMQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 346
DB 300 DKTTIILDAHTGEAKQOFPFHSAPALDVDMQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 359

QY 347 GHTNEVNAIKWDPGKGLLASCSDMTLKIWSMKQDCVHDLQAHKEIYTIKWSPTGCT 406
DB 360 GHTNEVNAIKWDPGKGLLASCSDMTLKIWSMKQDCVHDLQAHKEIYTIKWSPTGCT 419

QY 407 NNPNANMLASGFSDFSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGSPDKCV 466

DB 420 SNPNANMLASGFSDFSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGSPDKCV 479

QY 467 HIWNTQTGALVHSYRGTFGGIFECVMAAGKVGASASDGSVCVLDLURK 514
DB 480 HIWNTQSGSLVHSYRGTFGGIFECVMAAGKVGASASDGSVCVLDLURK 527

RESULT 6
Q8BMMO PRELIMINARY; PRT; 527 AA.
ID Q8BMMO;
AC Q8BMMO; (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Transducin.
GN TBL1X.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pituitary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
DR EMBL; AK030547; BAC27015.1; -;
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SMO0667; Lish; 1.
DR SMART; SMO0320; WD40; 8.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
SQ SEQUENCE 527 AA; 56832 MW; 0466F764ABA25CB0 CRC64;

Query Match 90.2%; Score 2462.5; DB 11; Length 527;
Best Local Similarity 87.5%; Pred. No. 1.6e-165;
Matches 462; Conservative 27; Mismatches 24; Indels 15; Gaps 3;

QY 1 MSISDEVNPLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQV 60
DB 1 MSITSDEVNPLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQV 60

QY 61 EAEVSINEDGTLFDGRIPIESLSLIDAVPDVQVOTROQAYRDKLAQQ--HAAAAAATAAT 118
DB 61 EAEISINEDGTVFDCRIPIESLSLIDAVPDVQVOTROQAFREKLTQQOANAAAAATAAT 120

QY 119 -----NQGSAGKNGENTANGENGATIANNHDTDMVEVDGVEIIPSKAVLVR 166
DB 121 ATSTAATTPAAAQQPPKNGEATVNGEANGAHAI--NNHSPKMEIDGVDVIPPSTATVLR 179

QY 167 GHESEVFCIAWNPVSDLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPSNK 226
DB 180 GHESEVFCIAWNPVSDLLASGSGDSTARINWLNENSGSTQLVLRHCIREGGHDVPSNK 239

QY 227 DVTSLDWSNSETLLATGSDGFARITWKGKNGLASTLQGHKGPFIKWKNGKNGIILSAGV 286
DB 240 DVTSLDWSNSETLLATGSDGFARITWKGKNGLASTLQGHKGPFIKWKNGKNGIILSAGV 299

QY 287 DKTTIILDAHTGEAKQOFPFHSAPALDVDMQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 346
DB 300 DKTTIILDAHTGEAKQOFPFHSAPALDVDMQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 359

QY 347 GHTNEVNAIKWDPGTGNLLASCDDMTLKISWKQDNCVHDLQAHNKETIYTIKWSPTGPGT 406
 DB 360 GHTNEVNAIKWDPFSGMLLASCDDMTLKISWKQDNCVHDLQAHNKETIYTIKWSPTGPGT 419
 QY 407 NNPANMLASAFSTSTVRLWVDVDRGICHTILTKHQEPVYSVAFSPDGRYLAGSGSFDKCV 466
 DB 420 SNPNSNMLASAFSTSTVRLWVDVDRGICHTILTKHQEPVYSVAFSPDGRYLAGSGSFDKCV 479
 QY 467 HWTQTGALVHSYRGTTGIFEVCNNAAGDKVGASASDGSVCVLDLRK 514
 DB 480 HWTQTGALVHSYRGTTGIFEVCNNAAGDKVGASASDGSVCVLDLRK 527

RESULT 7
 Q86UY2 PRELIMINARY; PRT; 577 AA.

AC Q86UY2;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE TBLIX protein.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.B., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.P., Casavant T.L., Schetz T.B.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman X., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 RA Rodriguez A.C., Grinchwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Strausberg R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL, BC052304; AAH52304.1; -.
 DR InterPro; IPR006594; LISH.
 DR InterPro; IPR001690; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PRO0320; GPROTEINBRPT.
 DR ProDom; PDOC0018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00882; WD_REPEATS_2; 6.
 DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
 DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
 SQ SEQUENCE 577 AA; 62495 MW; D830A37781E2A15C CRC64;

Query Match 89.3%; Score 2439; DB 4; Length 577;
 Best Local Similarity 86.0%; Pred. No. 8.2e-164;
 Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

QY 1 MSISSDENVLYRYLOESGFSAFTFGIESHSIQSNGINGALVPPAALISITOKGLQYV 60
 DB 52 MSITSDDENVLYRYLOESGFSAFTFGIESHSIQSNGINGALVPPAALISITOKGLQYV 111
 QY 61 EAEYSINEDGTFDGRPIESLSLIDAVMPDVVOTRQOAYRDKLAQQAQAAAAAATAA 118
 DB 112 EAEYSINEDGTFDGRPIESLSLIDAVMPDVVOTRQOAYRDKLAQQAQAAAAAATAA 171
 QY 119 -----NOQSAKNGENTANGENGAHTIANNHTDMMEVDGVEIPSNKAVVLRG 167
 DB 172 ATAATTTAGVSHONPSKREATVNGEENRAHSV-NNHAKPMEIDGVEIPSSKATVLRG 230
 QY 168 HSEVFICANVPVSDLLVSGSGDSTARINWLSNSTSGPTQLVLRHCRIRGGGDVPSNKG 227
 DB 231 HSEVFICANVPVSDLLVSGSGDSTARINWLSNSTSGPTQLVLRHCRIRGGGDVPSNKG 290
 QY 228 VTSLDNMSGTLLATGSDYGFARITWKDGNLASTLQGHKGPFIKLNKKNKGNFILSAGVD 287
 DB 291 VTSLDNMSGTLLATGSDYGFARITWKDGNLASTLQGHKGPFIKLNKKNKGNFILSAGVD 350
 QY 288 KTTIINDAHTGEAKQPPFHSAPALVDWQSNNTFASCSTDMCIHCKLQGDPIKTFQG 347
 DB 351 KTTIINDAHTGEAKQPPFHSAPALVDWQSNNTFASCSTDMCIHCKLQGDPIKTFQG 410
 QY 348 HTNEVNAIKWDPGTGNLLASCDDMTLKISWKQDNCVHDLQAHNKETIYTIKWSPTGPGT 407
 DB 411 HTNEVNAIKWDPGTGNLLASCDDMTLKISWKQDNCVHDLQAHNKETIYTIKWSPTGPGT 470
 QY 408 NPNANMLASAFSTSTVRLWVDVDRGICHTILTKHQEPVYSVAFSPDGRYLAGSGSFDKCV 467
 DB 471 NPNANMLASAFSTSTVRLWVDVDRGICHTILTKHQEPVYSVAFSPDGRYLAGSGSFDKCV 530
 QY 468 IWTQTGALVHSYRGTTGIFEVCNNAAGDKVGASASDGSVCVLDLRK 514
 DB 531 IWTQTGALVHSYRGTTGIFEVCNNAAGDKVGASASDGSVCVLDLRK 577

RESULT 8
 Q8COAL PRELIMINARY; PRT; 412 AA.

AC Q8COAL;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Transducin (Fragment).
 GN TBLIX.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 DR EMBL; AK031937; BAC27612.1; -.
 DR PIR; PT0651; PT0651.
 DR MGD; MGI:1336172; Tbl1x.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PRO0320; GPROTEINBRPT.
 DR ProDom; PDOC0018; WD40; 3.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00882; WD_REPEATS_2; 6.
 DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
 FT NON TER 1
 SQ SEQUENCE 412 AA; 44356 MW; EB78910B6D9E5237 CRC64;


```

179 ACNQTSGNSTSTPAGDLAAPCASQKQSQNSNEAGSSSGNAGNANATSTDDAASS 238
136 ---NG-----AHTIANHNTDM-----148
239 TSTNGNSTSSVVEQPTSGLTGAGTSTSTNPDAAAGGASTATGSKAPGAVTIRVGAQ 298
149 -----MEVDGVEIPSNKAVLRGHSEV 172
299 GNNVQSGSSNAASSAPSGTISSTSGAGTAAALVPMIDIDENIEIPESKARVLRGHSEV 358
173 FICAMPVSDLLVSGSDSTARIWLSNSTSGPTQLVLRHCITREGGQDVPFNKDVTSID 232
359 FICAMPNPSRDLASGSDSTARIWLSNSTSGPTQLVLRHCITREGGQDVPFNKDVTSID 417
233 WNSEGTLTATGSDGPARITWKGDNULASTLGQHGPIFALKWKNKGNFILLSAGVDTII 292
418 WNCDSGLLATGSDGYARIWKTGDRLASTLGQHGPIFALKWKNKGNFILLSAGVDTII 477
293 WDAHTGEAKQPPFHSAPALVDVQNNNTFASCSSTDCMHVCKLGQDRPIKTFQGHTEV 352
478 WDASTGCTQCFAPFHSAPALVDVQNNNTFASCSSTDCMHVCKLGQDRPIKTFQGHTEV 537
353 NAIKWPTGNTLLASCSDMTLKIWSKQNCVHDLOAHNKEIYTIKWSPTGPTNNPNAN 412
538 NAIKWPCQGLLASCSDMTLKIWSMNRDRCCHDLQAHSEIYTIKWSPTGPTNNPNNTN 597
413 LMLASAFSTVRLWVDVDRGICHTLTKEQFPVYSVAFSPDGRYLAGSFDKCVHWNTO 472
598 LILASAFSTVRLWVDVDRGICHTLTKEQFPVYSVAFSPDGRYLAGSFDKCVHWNTO 657
473 TQALVHSYRGTTGIFPCVWNAAGDKVGSASDGSVCVLDLRK 514
658 TQOLVHSYRGTTGIFPCVWNSXGTVKVGASDGSVCVLDLRK 699

RESULT 10
Q9FN19 PRELIMINARY; PRT; 613 AA.
ID AC Q9FN19
DC DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Genomic DNA, chromosome 5, TAC clone:K8K14 (AT5G67320/K8K14_4).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT Physically assigned pl clones."
RL DNA Res. 4:401-414(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang F.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

```

RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.W., Chang C.H., Dale J.M., Deng J.M., Hayashizaki Y.,
RA Hsuan V.W., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M.,
RA Southwick A., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yamada K., Yu G., Yuan S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007645; BAB09017.1; -
DR EMBL; AY057698; AAL15328.1; -
DR EMBL; AY143932; AAN28871.1; -
DR InterPro; IPR006594; Lish; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PDC00018; WD40; 2.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00882; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEAT_REGION; 1.
DR Repeat; WD repeat.
KW SEQUENCE 613 AA; 69772 MW; 04F4072311C0E76 CRC64;
Query Match 50.8%; Score 1388; DB 10; Length 613;
Best Local Similarity 44.2%; Pred. No. 1.4e-89;
Matches 272; Conservative 90; Mismatches 145; Indels 108; Gaps 7;
QY 2 SISDEVNLFVRYLOESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKLOYVE 61
DB 3 SLTSVELNLFVRYLOESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKLOYVE 62
QY 62 -----AEVSINEDGTLFDGPIESLSLDVAMPDVVQTOQAYRKLQAQHAHAAAA 114
DB 63 MEANLSNEVDIDEDFSFF--OPLDLISKVQKELQDLMLREKKERDEKEREKSKENDK 120
QY 115 AAATNQGSKAKNGENTANGE-----GDVE-----ENGAHTIAN 143
DB 121 GVEREHGDRNRAKORHEKQEREREREKEREKEREKEREKEREKEREKEREKEREIFE 180
QY 144 NHTDMEVD-----156
DB 181 REKRLKLEKEREIEREREREKEREKEREKEREKEREKEREKEREKEREKEREKEREIFE 240
QY 157 -----IPSNKAVVLRGHSEVFTCAWNPVSDLLVSGSDSTARIWLSNSTS 204
DB 241 DIVMTPTSGTSHIPNSDVRILEGHTSEVCACAWSPASLLASGSDATARIWLSPEGSFK 300
QY 205 GP-----TQLVLRHCIRREGQDVPFNKDVTSLDWNSGTLTATGSDGPARITWKGDN 258
DB 301 AVHTGRNIALILKHA--KXGSKNEKSDVTLTLDWNGEGTLTATGSDGQARIWTLNGL 357
QY 259 ASTLGQHGPIFALKWKNKGNFILLSAGVDTIIWDAHTGEAKQFPFHSAPALDVQNS 318
DB 358 ISTLSKHKGPFPFSLKWKGGDYLLTGSVDRTAVVMDVKAEWKQFPFHSAPALDVQNS 417
QY 319 NNTFASCSSTDCMHVCKLGQDRPIKTFQGHTEVNAIKWDPNTGNLLASCSDMTLKIWSM 378
DB 418 NVSFATSSDTSMIYLCIKIGETRPACTFTGHQEVNCKWDPNTGNLLASCSDMTLKIWSM 477
QY 379 KQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANMLASAFSTVRLWVDVDRGICHTL 438
DB 478 KQSTFVHDLREHTKEIYTIKWSPTGPTNNPNANMLASAFSTVRLWVDVDRGICHTL 537
QY 439 TKHQPVPVSAFSPDGRYLAGSFDKCVHWNTOGTALVHSYRGTTGIFPCVWNAAGDKV 498
DB 538 NGHREPVYSLAFSPNGEYLAGSFDKCVHWNTOGTALVHSYRGTTGIFPCVWNAAGDKV 597
QY 499 GASASDGSVCVLDLR 513
```

Db 598 AACFADNSVCVLDLR 612

RESULT 11

Q8VEG3 PRELIMINARY; PRT; 201 AA.
 AC Q8VEG3;
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Similar to Iral protein (fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC018512; AAH18512.1; -;
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 4.
 DR ProDom; PD000018; WD40; 2.
 DR SMART; SMC0320; WD40; 4.
 DR PRINTS; PR00320; GPROTEINERPT.
 DR PROSITE; PS00678; WD_REPEATS_1; 2.
 DR PROSITE; PS00082; WD_REPEATS_2; 3.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT NON_CTER
 SQ SEQUENCE 201 AA; 22038 MW; 6B945F137B491818 CRC64;

Query Match 40.3%; Score 1101; DB 11; Length 201;
 Best Local Similarity 99.5%; Pred. No. 5.4e-70;
 Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 314 VDMQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPGTGNLLASCSDMTL 373
 Db 1 VDMQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPGTGNLLASCSDMTL 60
 QY 374 KISMKQDCNCVHDLQAHNKEIYTIKWSPTGPTGNPNANMLASAFDSTVLWDVDRGI 433
 Db 61 KISMKQDCNCVHDLQAHNKEIYTIKWSPTGPTGNPNANMLASAFDSTVLWDVDRGI 120
 QY 434 CIHTLKHQEPVSVAFSPDGRYLAGSPDKCVHIWNTQTGALVHSYRGCTGGEFVCWNA 493
 Db 121 CIHTLKHQEPVSVAFSPDGRYLAGSPDKCVHIWNTQTGALVHSYRGCTGGEFVCWNA 180
 QY 494 AGDKVGASADSGSVCLDLRK 514
 Db 181 AGDKVGASADSGSVCLDLRK 201

RESULT 12

Q95RJ9 PRELIMINARY; PRT; 524 AA.
 ID Q95RJ9;
 AC Q95RJ9;
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE L24373p.
 GN EBI OR CG4063.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Berkeley;
 RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Chame M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nunco J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY061326; AAL28874.1; -;
 DR FlyBase; FBgn0023444; ebi;
 DR GO; GO:000074; P:regulation of cell cycle; IMP.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001005; MYB_DNA_binding.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 3.
 DR PRINTS; PR00320; GPROTEINERPT.
 DR ProDom; PD000018; WD40; 1.
 DR SMART; SMC0667; Lish; 1.
 DR SMART; SMC0320; WD40; 3.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00037; MYB 1; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 2.
 DR PROSITE; PS00082; WD_REPEATS_2; 3.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 524 AA; 52529 MW; 6C3D86110BA18D65 CRC64;

Query Match 33.7%; Score 920.5; DB 5; Length 524;
 Best Local Similarity 40.3%; Pred. No. 1.2e-56;
 Matches 217; Conservative 37; Mismatches 67; Indels 217; Gaps 12;
 QY 1 MSISDEVNFIYRYLOESGFHSQFTGIESHSQSNINGALVPPAALISIIQKGLQYV 60
 Db 1 MSFSDEVNFIYRYLOESGFHSQFTGIESHSQSNINGALVPPAALISIIQKGLLYT 60
 QY 61 EAEVSIINEDGTLFDGRIEISLIDAVMPDV-----VQTRQ----- 96
 Db 61 EAEVSIINEDGTLFDGRIEISLIDAVMPDV-----VQTRQ----- 96
 QY 97 -----QAYRDKLA-----QQHAAAAAATAATNOGSAKNGENT 130
 Db 119 NAKPEIKIEPGTVAGSAGGKNIAGSTGTSTPTDQSAEVDSSGNAANNAGGTIAGNG 178
 QY 131 ANGE----- 135
 Db 179 AGNQASTGSGNSTSTPAGDGLAAGASQKQSNSEAGSSSGNAGNANATSTDDRASS 238
 QY 136 -----AHTIANNHTDM----- 148
 Db 239 TSTNGNSTSSSVQPTSGLTGAGTVSTSNPDAAAAGSGASTATGSKAPSGAVTIRVGAQ 298
 QY 149 -----NEVDGDEIIPSKAVLRGHESEV 172
 Db 299 GNNVQSGSSNAQSAPSGTSSSTSGGAGTFAALVPMIDENIEIPESKARVLRGHESEV 358
 QY 173 FICAWNPSVLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPFNNKDVTSLD 232
 Db 359 FICAWNPSRLLASGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPFNNKDVTSLD 417
 QY 233 WNSEGTLATGSGDGFARIWTKDGNLASTLGQHKGPFAKKNKGNFILLSAGVDKTTII 292
 Db 418 WNCDSLLATGSGDGFARIWTKDGNLASTLGQHKGPFAKKNKGNFILLSAGVDKTTII 477
 QY 293 WDAHTGEAKQFPFHSAPALDQSNNTFASC-----STDMCIHV-CK 335
 Db 478 WDAHTG-----FNEPAICL-----SQSSSLGCLADKPGCLLLQYGSADTCVPVGCK 524

RESULT 13

Q8X1P4
 ID Q8X1P4
 AC Q8X1P4;
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Beta transducin-like protein HET-E2C*4.
 GN HET-E.
 OS Podospora anserina.

Search completed: August 9, 2004, 16:48:11
Job time : 35.6667 secs